

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Transient transcriptome sequencing:
development and applications in human cells

Margaux Michel
aus
Villeneuve-la-Garenne, Frankreich
2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 17.07.2016

Margaux Michel

Dissertation eingereicht am 21.07.2016

1. Gutachter: Prof. Dr. Patrick Cramer

2. Gutachter: PD Dr. Dietmar Martin

Mündliche Prüfung am 05.09.2016

Summary

Transcription is one of the most important processes in life. Correct timing and regulation of transcription is responsible for cellular development, identity, adaptivity to environmental cues, differentiation and many other processes. Dysregulation of transcription can lead to the development of cancer and other diseases. The transcriptome of polymerase (Pol) II consists of messenger RNA (mRNA) encoding proteins and non-coding RNA (ncRNA). Due to their short half-life ncRNAs are not easy to catch and thus, many of their functions remain elusive. We set out to develop an appropriate method in human cell lines to learn more about ncRNAs, their occurrence, regulation, and kinetics.

To this end we developed transient-transcriptome sequencing (TT-seq) to detect and map transient full-length RNAs *in vivo*. The method is based on 4-thiouridine (4sU) sequencing, where a uridine analog is supplied to cells and incorporated into nascent RNA. Fragmentation of the RNA prior to isolation of labeled RNA is the key component of the TT-seq method. This step leads to a drastic enrichment of newly-synthesized RNAs and enables detection of unstable RNAs. We employed TT-seq in the human leukemic cell line K562 and could detect and map thousands of intronic RNAs, ncRNA classes such as enhancer RNAs (eRNAs) and RNA downstream of the polyadenylation (pA) site. TT-seq enabled the analysis of synthesis and degradation rates by comparing the total RNA pool to the nascent labeled RNA fraction. In particular, we found that ncRNAs especially eRNAs are short-lived with a median half-life of 2 minutes. Analysis of RNAs downstream of the pA site enabled us to map human transcription termination sites. We found that termination sites are enriched for a (C/G)₍₂₋₆₎A kmer followed by a (T/A)₍₃₋₆₎ kmer. This nucleotide composition can favor backtracking and/or pausing due to the change in the energy landscape of the RNA:DNA hybrid within Pol II. We propose that this facilitates termination by the exonuclease Xrn2 and is dependent on the presence of a pA site.

To further understand eRNAs and their synthesis kinetics, we performed TT-seq in Jurkat T cells during the first 15 min of T cell activation. We found that thousands of mRNAs and ncRNAs transcripts were differentially expressed during the time course in the TT-seq samples. Interestingly, not a single transcript was significantly differentially expressed in the Total RNA-seq samples. This indicates that the sensitivity of TT-seq can also be used to revisit known pathways. We found that the expression levels as well as their change over time of the eRNA and the paired mRNA are highly correlated. In contrast to previous reports (Arner et al, 2015), showing that eRNA expression precedes its paired mRNA expression, we found simultaneous activation or downregulation of mRNA and eRNA pairs that are differentially expressed after 15 min.

Taken together, we developed the new method TT-seq that enables mapping and analysis of newly synthesized RNAs and can determine synthesis and degradation rates. In addition, the new method is especially sensitive for transient RNAs and allows detection of very rapid expression changes. This method can be applied to any organism, which is able of 4sU (or 4-thiouracil; 4tU) uptake. Therefore, it can be broadly used to investigate many fascinating outstanding questions such as the mechanism of Pol II termination, ncRNAs degradation pathways or eRNA role in promoter activation.

Publications

Parts of this thesis have been published or are in the process of being published:

2016 **Enhancer RNAs and their paired promoter RNAs arise simultaneously after stimulation**

M. Michel*, C. Demel*, K. Hofmann, A. Sawicka, B. Zacher, B. Schwalb, S. Krebs, H. Blum, J. Gagneur, and P. Cramer (* joint first authorship)

Manuscript in preparation.

Author contributions: MM carried out all experiments except for the Pol II phosphoisoforms (Ser5P, Ser2P, Tyr1P) ChIP-seq experiments performed by KH. CD performed all bioinformatical analyses. AS performed the enhancer promoter pair confirmation experiment. BZ provided scripts and assistance with the transcriptome annotation. BS provided scripts. SK performed all sequencing reactions, supervised by HB. JG and PC designed and supervised research. MM, CD, JG, and PC prepared the manuscript, with input from all authors.

2016 **Accurate promoter and enhancer identification in 127 ENCODE and Roadmap Epigenomics cell types and tissues by GenoSTAN**

B. Zacher, **M. Michel**, B. Schwalb, P. Cramer, A. Tresch, and J. Gagneur

Under revision.

Author contributions: BZ, JG and AT developed the statistical methods and computational workflow of the study. BZ developed and implemented all software and scripts and carried out all computational analyses. BS helped with preprocessing of the K562 data set. MM and PC helped with interpretation of the biological results. BZ, JG and AT wrote the manuscript with input from all authors.

2016 **TT-seq maps the human transient transcriptome**

B. Schwalb*, **M. Michel***, B. Zacher*, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer (* joint first authorship)

Science, 352(6290), 1225-1228.

Author contributions: MM designed and carried out all experiments. BS carried out all bioinformatics analysis except transcript calling, RNA classification, analysis of U1 sequence motifs, and prediction of RNA secondary structure, which were carried out by BZ. BZ, JG and AT developed the chromatin state annotation. KF designed RNA spike-in probes. CD established the spike-in normalization method. BS, JG and PC designed research. JG and PC supervised research. BS, MM, JG, and PC prepared the manuscript, with input from all authors.

2013 **Transitions for regulating early transcription**

M. Michel and P. Cramer

Cell, 153(5), 943-944.

Author contributions: MM and PC prepared the manuscript.

Contents

Erklärung	3
Eidesstattliche Versicherung	3
Summary	5
Publications	7
Contents.....	8
1 Introduction.....	11
1.1 RNA polymerase II transcription cycle	11
1.1.1 Transcription initiation	12
1.1.2 Transcription elongation.....	14
1.1.3 Transcription termination	14
1.1.4 Recycling of Pol II	15
1.2 Transient RNAs.....	16
1.2.1 ncRNAs	16
1.2.2 RNAs downstream of the pA site.....	16
1.2.3 eRNAs	17
1.3 RNA degradation pathways.....	17
1.3.1 Nuclear degradation pathways	17
1.3.2 Cytoplasmic degradation pathways.....	18
1.4 Methods to analyze RNA metabolism.....	19
1.4.1 Assessing synthesis and degradation rates	19
1.4.2 Analyzing transient RNAs.....	20
1.5 Aims of this thesis	23
2 Materials and Methods.....	25
2.1 Materials.....	25
2.1.1 Bacterial strain.....	25
2.1.2 Human cell lines	25
2.1.3 Media and supplements	25
2.1.4 Spike-ins	26
2.1.5 Primers.....	26
2.1.6 Thermal cycler programs.....	30
2.1.7 Plasmids.....	30
2.1.8 Buffers and solutions	31

2.1.9	Antibodies	32
2.2	Experimental methods	32
2.2.1	Human cell culture	32
2.2.2	Generation of spike-ins.....	33
2.2.3	TT-seq	34
2.2.4	ChIP-seq	39
2.2.5	<i>In vivo</i> termination experiment.....	42
3	Results and Discussion	45
3.1	TT-seq maps the human transcriptome	45
3.2	Enhancer RNAs and their paired promoter RNAs arise simultaneously after stimulation ...	65
4	Future Perspectives	79
4.1	ncRNA surveillance pathway in humans	79
4.2	Different underlying mechanisms leading to promoter-proximal peaks of Pol II.....	79
4.3	Splicing rate.....	80
4.4	Initiation and synthesis rate.....	81
4.5	Pol II CTD modifications at ncRNAs loci	81
4.6	eRNA function	81
4.7	Concluding remarks	83
5	References.....	84
6	Abbreviations	98
	Acknowledgments	99

1 Introduction

Transcription is a one of the most fundamental biological processes. The genetic information is encoded in the DNA, which is transcribed to RNA that serves as a template for protein synthesis. This is well known as the central dogma of molecular biology (Crick, 1970). Transcription is performed by DNA-dependent RNA polymerases (Pol; Roeder & Rutter, 1969). Eukaryotes, except plants, have three nuclear Pols: Pol I, Pol II and Pol III (Cramer et al, 2008). Pol I transcribes ribosomal RNAs (rRNAs), Pol II messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs) such as long non-coding RNAs (lncRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) and Pol III transcribes transfer RNA (tRNAs) and 5S rRNAs (Zylber & Penman, 1971). Plants have two additional non-essential nuclear Pols: Pol IV and Pol V that synthesize small interfering RNAs (siRNAs; Herr et al, 2005; Ream et al, 2009). Furthermore, eukaryotes encode for a mitochondrial RNA polymerase active in mitochondria (Ringel et al, 2011).

All nuclear polymerases have a conserved 10-subunit core and additional peripheral subunits (Cramer et al, 2008). Pol II is a 12-subunit protein complex of 512 kDa that consists of the subunits Rpb1 to Rpb12 (Armache et al, 2003). During transcription, Pols and in particular Pol II associate with a plethora of proteins such as general transcription factors and other factors that regulate Pols' activity and ensure the correct course of events.

1.1 RNA polymerase II transcription cycle

The process of transcription can be delimited into 4 phases: initiation, elongation, termination and recycling (Figure 1; Shandilya & Roberts, 2012). A unique property of Pol II is the large Rpb1 subunit with its extended C-terminal domain (CTD; Cramer, 2004). The CTD consists of a number of heptapeptide repeats, 26 in *Saccharomyces cerevisiae* and 52 in human. The majority of the heptapeptides are highly conserved and are composed of Tyrosine (Tyr), Serine (Ser), Threonine (Thr) and Proline (Pro) in the order Tyr1, Ser2, Pro3, Thr4, Ser5, Pro6 and Ser7 (Corden et al, 1985). All amino acids, except Pro, are capable of phosphorylation in a dynamic manner during different phases of the transcription cycle, which results in the so termed CTD code (Egloff & Murphy, 2008). Furthermore, Ser and Thr can be glycosylated and Pro can either be in cis or trans conformation (isomerized; Eick & Geyer, 2013; Zeidan & Hart, 2010). Each phase of the transcription cycle is characterized by

a specific CTD modification pattern that serves as a binding platform for multiple factors (Buratowski, 2009; Heidemann et al, 2013).

1.1.1 Transcription initiation

Transcription by Pol II starts by assembling the pre-initiation complex on the DNA (Figure 1; Cheung et al, 2011). This is achieved by cooperational binding of general transcription factors TFIID, TFIIA, TFIIB, TFIIF and TFIIIE and the recruitment of Pol II (Hahn, 2004; Sainsbury et al, 2015; Thomas & Chiang, 2006). The pre-initiation complex can be further stabilized and enhanced by the Mediator complex or other co-activators such as SAGA (Chen et al, 2012; Svejstrup et al, 1997). Subsequently, 11 to 15 base pairs (bp) DNA around the transcription start site (TSS) are melted and the template strand is positioned in the active site cleft of Pol II to form the Open Complex (Wang et al, 1992). During initiation, the cyclin-dependent kinase (Cdk) 7, which is part of the TFIIH complex, phosphorylates Ser5, Ser7 and Tyr1 of the CTD (Komarnitsky et al, 2000). The mediator complex subunit Cdk8 has also been shown to phosphorylate Ser5 (Liu et al, 2004). Once the transcription bubble is positioned within the active site, cycles of abortive transcription of 3 to 10 nucleotides (nt) of RNA ensue (Holstege et al, 1997). Once the transcribed RNA reaches a length of 30 nt, the polymerase is able to escape the promoter, termed promoter clearance (Luse, 2013). By then, the capping enzyme is recruited to the initiation complex via Ser5 phosphorylation (Ser5P; Lidschreiber et al, 2013). The capping enzyme links the 5'-end of the RNA to a 7-methylguanosine via a 5'-5' triphosphate bridge (Wei et al, 1975; Wei & Moss, 1977). The 5' cap protects the RNA from 5' exonucleolytic cleavage and promotes translation in the cytoplasm (Schwer et al, 1998; Sonenberg & Hinnebusch, 2009).

1.1.1.1 Promoter-proximal pausing

In higher eukaryotes, transcription can come to a halt 30 to 150 bp after the TSS on particular genes during what is termed promoter-proximal pausing (PPP; Figure 1; Adelman & Lis, 2012; Core et al, 2008b). PPP has been mainly defined by an occurrence of a Pol II peak in chromatin immuno-precipitation (ChIP) experiments near the TSS (Zeitlinger et al, 2007). However, ChIP experiments have to be interpreted with caution since a peak in a profile can be explained by: first, a large amount of crosslinked proteins at this position or second, the long residing time of a few target proteins at the same position or third, a mixture of both. TFIIH inhibition experiments indicated that the majority of polymerases observed as paused

are in fact undergoing repeated rounds of termination and re-initiation, whereas only a small fraction of genes have stably paused Pol II (hour-range) near the TSS (Chen et al, 2015). Furthermore, accumulation of Pol II near the TSS has also been shown to be a consequence of different initiation and elongation speeds without any apparent regulatory functions (Ehrensberger et al, 2013).

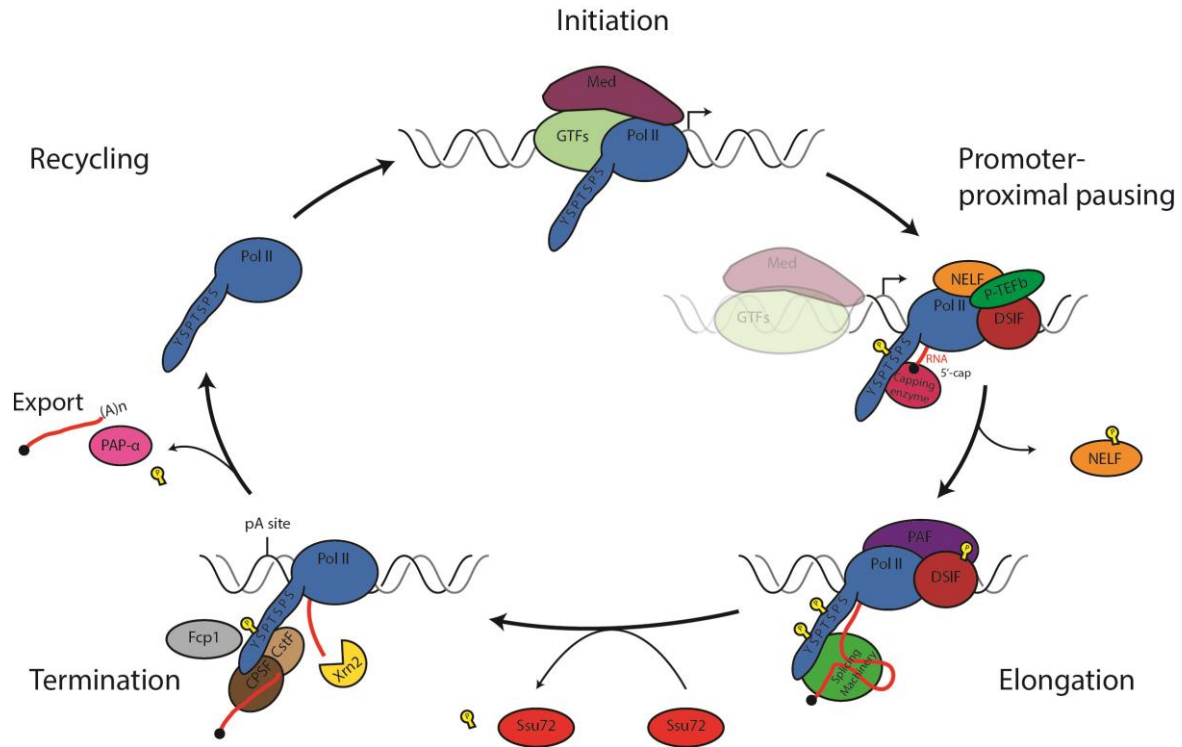


Figure 1: Schematic of the human transcription cycle. During the initiation phase, polymerase II (Pol II) is recruited to the promoter via general transcription factors (GTFs), this interaction is enhanced and stabilized by Mediator. After initiation of transcription and Serine 5 phosphorylation (Ser5P), the nascent RNA is capped by the capping enzyme. Recruitment of the capping enzyme is mediated by Ser5P. Pol II can pause downstream of the transcription start site, termed promoter-proximal pausing. The negative elongation factor (NELF) and the DRB sensitivity inducing factor (DSIF) are responsible for Pol II pausing. The kinase positive transcription elongation factor (P-TEFb) phosphorylates NELF, DSIF and the C-terminal domain (CTD) of Pol II at Serine 2 (Ser2P). This releases Pol II from promoter-proximal pausing. Transcription elongation ensues. During this phase, splicing factors are recruited to the CTD via Ser5P/Ser2P and splice the nascent RNA. The phosphatase Ssu72 binds transiently to the CTD and dephosphorylates Ser5P. The last phase on the template DNA is the termination phase. Termination factors CPSF and CstF are recruited to the CTD via Ser2P and cleave the RNA at the polyadenylation (pA) site. Subsequently, Ser2P is dephosphorylated by Fcp1. Last, Pol II is terminated by Xrn2 and recycled for the next transcription cycle. The RNA is polyadenylated by the polyA polymerase (PAP-α) and exported to the cytoplasm.

PPP is promoted by NELF and DSIF (Gilchrist et al, 2008). Release of Pol II into productive elongation is mediated by Cdk9 that is part of the positive transcription elongation factor (P-TEFb; Bieniasz et al, 1999). Polymerase release is realized by phosphorylation of the CTD at

Ser2 (Ser2P), NELF and DSIF via the P-TEFb complex (Peterlin & Price, 2006). In addition, Cdk12, Cdk13 and BRD4 have also been reported as Ser2 kinases (Bartkowiak et al, 2010; Devaiah et al, 2012). After phosphorylation, NELF disassociates from Pol II and thereon DSIF acts as a positive elongation factor. Many functions have been attributed to PPP since its discovery in *Drosophila melanogaster*: synchronous activation of genes during development, retaining an open chromatin landscape at promoters, faster activation of genes or checkpoint for correct mRNA capping (Gilmour & Lis, 1986; Nechaev & Adelman, 2011).

1.1.2 Transcription elongation

Once Ser2 of the CTD has been phosphorylated and Pol II has cleared away from the promoter-proximal region, productive elongation can ensue and is supported by elongation factors (Figure 1). Together with Pol II, elongation factors form a stable multi-component elongation complex (Mayer et al, 2010). The elongation complex can harbor DSIF, TFIIS, the PAF complex, elongin, FACT and other proteins, which all together ensure maximal processivity of Pol II (Kettenberger et al, 2004; Martinez-Rucobo et al, 2011; Shandilya & Roberts, 2012). Ser2P serves as a binding platform for splicing factors that travel with Pol II and splice introns during transcription elongation (Gu et al, 2013). Gradually over the gene body, the S5P mark is removed by phosphatases such as Ssu72 (Krishnamurthy et al, 2004). Ssu72 also removes Ser7 phosphorylation (S7P) on mRNA genes (Zhang et al, 2012).

1.1.3 Transcription termination

Termination of protein-coding genes is polyadenylation (pA) signal dependent (Proudfoot, 2011). Ser2P acts as a binding platform for termination factors (Figure 1; Gu et al, 2013). The pA signal is a well-defined RNA sequence 5'-AAUAAA-3' followed by a G/U rich stretch (Gil & Proudfoot, 1984). Once the pA signal is transcribed and emerges in the nascent transcript, it is recognized by termination factors such as CPSF leading to Pol II slowdown or pausing (Mischo & Proudfoot, 2013). The emerging G/U rich stretch is recognized by CstF, which leads to cleavage of the RNA by CPSF 10 to 30 nt downstream of its binding site (Kuehner et al, 2011). 3' processing of the mRNA involves addition of a polyA-tail of approximately 250 nt through the polyadenylation polymerase PAP- α (Balbo & Bohm, 2007). The mRNA is then exported to the cytoplasm, where it can be translated.

Not much is known about termination of ncRNAs in metazoans (Porrua & Libri, 2015). So far, the only alternative pathway is the termination of snRNAs by the Integrator complex via

Ser7P (Egloff et al, 2007; Egloff et al, 2010; Skaar et al, 2015). The Integrator complex contains homologs of the CPSF subunits (Baillat et al, 2005). Although Pol II shows phosphorylation of Ser7 during transcription of all transcript classes, disruption of this phosphorylation only shows an effect on expression of snRNA transcripts (Egloff et al, 2007).

Termination involves not only RNA release from Pol II but also Pol II release from the DNA template. There are two models for termination described: the allosteric and the torpedo model. Several studies hint to a combination of both models *in vivo*. In the first model, Ser2P is dephosphorylated by Fcp1 after the pA site (Cho et al, 2001). Loss of Ser2P causes loss of elongation factors after the pA site. This leads to a destabilization and/or conformational change of the ternary complex and renders Pol II less processive, with lower synthesis rate and leads to Pol II pausing or dissociation from the template (Keene et al, 1999; Logan et al, 1987). The second model is the torpedo model (Connelly & Manley, 1988), where cleavage of the RNA at the pA site leads to an unprotected 5' RNA end, which is recognized by a 5'-3' exonuclease. The exonuclease's processivity rate competes with Pol II's synthesis rate, eventually catches up with Pol II and dislodges it from the DNA template. So far, the torpedo model had only been shown for the yeast protein Rat1 (Kim et al, 2004). However, recent findings confirmed that the torpedo model also holds true in human cells (Fong et al, 2015). First, dominant-negative Xrn2 mutants in the active site showed termination defects genome-wide and second, Pol II elongation rate mutants shifted termination sites in comparison to wild type (Fong et al, 2015). Many aspects of the termination pathway remain elusive such as exact termination sites, nucleotide composition of termination sites and why are those favorable to termination.

1.1.4 Recycling of Pol II

Pol II is released in a hypo-phosphorylated state from the DNA template and thus, competent for re-initiation (Figure 1). Re-initiation is facilitated first, by gene bookmarking, where general transcription factors remain bound to the promoter during the transcription cycle (Sarge & Park-Sarge, 2005), and second, by gene looping, where the 5' end of a gene interacts through protein-protein interaction with the 3' end of the same gene (Calvo & Manley, 2003). Gene looping is promoted by phosphorylated TFIIB that directly interacts with CstF in yeast (Wang et al, 2010).

1.2 Transient RNAs

The most studied RNAs in living organisms are usually the ones that are the most stable and therefore easy to purify and analyze. RNAs that only live for few minutes are very hard to catch. Therefore, for a long time only protein-coding transcripts were assumed to be transcribed by Pol II and that the transcriptome represented only a very small percentage of the genome. Recent advances in methodologies and technologies, primarily the development of next-generation sequencing, have revoked this concept and showed that 90% of the human genome is transcribed (Wang et al, 2009; Wilhelm et al, 2008). This uncovered a multitude of ncRNAs with many different crucial functions for cell development and identity (Djebali et al, 2012; Eddy, 2001).

1.2.1 ncRNAs

Many ncRNAs have been mapped to this day (Andersson et al, 2014; Core et al, 2008b; Harrow et al, 2012). However, the function or rather functionality of many ncRNAs is far from being understood (Costa, 2010). ncRNAs can be long (~10 kb; lncRNAs) or short (~1 to 5 kb). Short ncRNAs can emerge from promoter regions or enhancer regions (eRNAs). Furthermore, short ncRNAs are dispersed in the genome, such as upstream of the TSS of a protein-coding gene (mRNA), on the opposite strand of an mRNA or far away from any annotated feature (Knowling & Morris, 2011). ncRNAs have been associated with many functions (Cawley et al, 2004). The MYCNOS ncRNA can silence protein-coding RNA through complementary base pairing and recruitment of the RNAi silencing machinery (Vadie et al, 2015). The ncRNAs Air and Xist can recruit chromatin modifier enzymes and regulate gene expression (Navarro & Avner, 2010). The 7SK ncRNA can sequester proteins such as P-TEFb (Diribarne & Bensaude, 2009).

1.2.2 RNAs downstream of the pA site

As previously mentioned, when Pol II transcribes over the pA site, the mRNA is cleaved. Nevertheless, Pol II is still associated with the DNA template and continues transcription. This leads to a new RNA with a new 5' end, which is uncapped (Proudfoot, 2011). It was shown for specific loci and genome-wide that the RNA downstream from the pA could be very long (Core et al, 2008b; Lian et al, 2008; Proudfoot, 1989). The exact mechanism of termination and release of Pol II remains elusive. A more detailed analysis of the transient RNAs, especially their length, 3' end and half-life could help understand termination.

1.2.3 eRNAs

eRNAs emerge from enhancers (De Santa et al, 2010a; Kim et al, 2010b). Enhancers are regulatory DNA regions that are bound by transcription and activation factors. These factors recruit the Mediator complex (Fan et al, 2006), which in turn binds and stabilizes the pre-initiation complex on a given promoter (Malik & Roeder, 2005). Enhancers have an enhancing effect on the transcription of a promoter they regulate (Banerji et al, 1981). Protein-protein interaction between activators, Mediator and the pre-initiation complex promotes long-range interaction via chromatin looping (Allen & Taatjes, 2015; Plaschka et al, 2015). This brings the enhancer and the promoter in close proximity even if they are located kilobases apart on the same chromosome (Petrascheck et al, 2005). Further proteins such as CTCF and cohesin support looping and delimit topologically associated domains (Hadjur et al, 2009; Ong & Corces, 2014; Parelho et al, 2008). It is not known whether eRNAs have a function or are the mere product of a high local concentration of initiation factors, transcription factors and Pols (Li et al, 2016). eRNAs have been associated with stabilization of enhancer-promoter looping (Hsieh et al, 2014; Li et al, 2013), but this was not true for every loci analyzed (Hah et al, 2013; Schaukowitch et al, 2014). Knockdown of eRNAs was shown in some examples to reduce target promoter expression (Li et al, 2013; Schaukowitch et al, 2014). In conclusion, further experiments have to be conducted to elucidate eRNA occurrences and functions in a cell. Development of a new method to analyze eRNAs in their full length without cellular perturbation would shed light on this matter.

1.3 RNA degradation pathways

1.3.1 Nuclear degradation pathways

RNA degradation is conducted by RNases that can be divided into three categories: endonucleases, 5'-3' exonucleases such as Xrn2 or 3'-5' exonucleases such as the exosome complex. Nuclear degradation pathways of Pol II transcripts include first, quality control of the mRNA and second, degradation of ncRNAs (Houseley & Tollervey, 2009).

1.3.1.1 Nuclear mRNA degradation

Quality control of mRNA requires that all processing steps are correctly completed before the RNA can be exported to the cytoplasm. This involves correct capping, splicing and polyadenylation. At each step nucleases can degrade the RNA, for example the 5'-3'

exonuclease Xrn2 at non-properly capped mRNAs, intron debranching by DBR1 and degradation by exonucleases for splicing errors and last, degradation also by Xrn2 of the RNA downstream of the pA site (Chapman & Boeke, 1991; West et al, 2008). Properly processed mRNAs are coated with RNA-binding factors that signalize that the mRNA is export-ready (Rougemaille et al, 2008). In yeast, the nuclear quality control complex TRAMP 3'-polyadenylates transcripts that failed quality control and those are subsequently degraded by the nuclear exosome complex (Houseley et al, 2006). In human cells, although homologs of the TRAMP complex components are present, functionality has not yet been reported (Houseley & Tollervey, 2009).

1.3.1.2 Nuclear ncRNAs degradation

In yeast, the Nrd1-Nab3-Sen1 complex mediates degradation of many ncRNAs such as snRNAs, snoRNAs, cryptic unstable transcripts, a few stable unannotated transcripts and all Nrd1-terminated transcripts (Schulz et al, 2013; Thiebaut et al, 2006; Xu et al, 2009). The Nrd1-Nab3-Sen1 complex is recruited to transcribing Pol II by binding to Ser5P of the CTD and to specific RNAs through a RNA-binding domain (Schulz et al, 2013; Steinmetz & Brow, 1996; Vasiljeva et al, 2008). The RNA is then targeted by the TRAMP complex and the exosome (Vasiljeva & Buratowski, 2006).

In humans, several ncRNAs, such as upstream antisense RNAs are targeted by the exosome complex (Preker et al, 2008; Schmid & Jensen, 2013). Recruitment of the exosome to specific RNAs is mediated by the NEXT complex that interacts with RNA through its subunit RBM7 that harbors a RNA binding site (Lubas et al, 2015). TSS-associated transcripts emerging from stalled Pol II are targeted by Xrn2 (Seila et al, 2008; Valen et al, 2011). ncRNA degradation pathways and mechanisms in human cells should be further explored in future studies.

1.3.2 Cytoplasmic degradation pathways

Once exported to the cytoplasm, the mRNA is subjected to a pioneer round of translation that is supported by the cap-binding complex and is important for quality control (Maquat et al, 2010). One of the quality controls assessed during the pioneer round of translation is correct splicing of the mRNA. The splicing machinery deposits an exon junction complex on newly ligated exon-exon junctions (Tange et al, 2004), which is assessed by the Nonsense-mediated decay machinery (Stalder & Muhlemann, 2008). In case of defective splicing and occurrence

of a premature stop codon before the last exon-exon junctions, the Nonsense-mediated decay targets the RNA for degradation (Houseley & Tollervey, 2009).

Other quality control pathways involve the Non-stop decay and the No-go decay. Non-stop decay occurs when an mRNA lacks a stop codon. As a result the ribosome cannot be released and is therefore stalled (Klauer & van Hoof, 2012). Non-stop decay is mediated by the Ski complex, which recruits the cytoplasmic exosome (Halbach et al, 2013; Schaeffer & van Hoof, 2011). No-go decay is a pathway also dealing with stalled ribosomes but in contrast to Non-stop decay, these are stalled on the transcript - for example due to strong secondary structures - and not at the 3'end; this leads to an endonucleolytic cleavage of the mRNA (Doma & Parker, 2006).

Steady-state rounds of translation follow the pioneer round of translation and are also supported by the cap-binding complex, which binds and recruits the translation initiation factor eIF4E (Isken & Maquat, 2008; Maquat et al, 2010). During translation, the mRNA is gradually deadenylated by Ccr4 and Caf1 from the Ccr4-NOT complex (Eulalio et al, 2009; Schwede et al, 2008). After Ccr4-NOT removal of the polyA-tail, the mRNA is decapped by Dcp1 and Dcp2 (Deshmukh et al, 2008), which leads to either 5' degradation by Xrn1 or 3' degradation by the exosome complex (Bonneau et al, 2009; Makino et al, 2015).

1.4 Methods to analyze RNA metabolism

1.4.1 Assessing synthesis and degradation rates

Each RNA has a specific synthesis and degradation rate that depends on its function in the cell and both rates define the amount of a given RNA in a cell (Wang et al, 2002). Both rates can vary due to external factors, cell cycle progression (Eser et al, 2014), hormonal stimuli (Ross, 1995), environment changes or cellular differentiation (Amorim et al, 2010; Jack & Wabl, 1988). Those changes will translate into changes in synthesis and/or degradation rates (Hargrove & Schmidt, 1989). These processes are therefore extremely important for RNA and protein levels and cellular functions (Perez-Ortin, 2007; Perez-Ortin et al, 2007).

In vivo determination of synthesis and degradation rates can be achieved by different experimental methods. Degradation rates can be assessed by specific transcriptional shut-off assays using thermal Pol II mutants (Herrick et al, 1990; Nonet et al, 1987), or specific

transcription inhibitors such as thiolutin and rifampicin (Bernstein et al, 2002; Grigull et al, 2004; Narsai et al, 2007). Synthesis rates can be determined for example by fluorescence microscopy assays (Darzacq et al, 2007), or by genomic run-on assays (Garcia-Martinez et al, 2004).

Another more powerful method is to use metabolic labeling since this enables estimation of both synthesis and degradation rates *in vivo* in a non-perturbing setup (Miller et al, 2011; Schwalb et al, 2012; Sun et al, 2012). The cells are fed with a nucleoside analog that is incorporated into nascent RNAs (Dolken et al, 2008). The labeled RNA can then be extracted from the Total RNA pool by affinity purification (Cleary et al, 2005), and analyzed by sequencing (Schulz et al, 2013). Mathematical modeling of the ratio of Total and Labeled RNA fractions enables determination of synthesis and degradation rates (Schwalb et al, 2012).

1.4.2 Analyzing transient RNAs

Nuclear transient RNAs can be analyzed through different experimental setups. The most common way is to deplete cellular factors responsible for degradation of transient RNAs, hence making the transient RNAs indirectly more stable for analysis. Factors can include exosome components (Carneiro et al, 2007; Preker et al, 2008) or Xrn2 (West et al, 2004). One of the downside of these methods is that cellular perturbation can lead to global or off-target effects (Sun et al, 2013).

Nuclear run-on assays such as PRO-seq/PRO-cap (Kwak et al, 2013), GRO-seq (Core et al, 2008b) or GRO-cap (Core et al, 2014), are one of the most sensitive methods when it comes to transient RNAs. Downside of these methods is that they are rather experimentally complicated and per se *in vitro*, since one isolates nuclei and resumes transcription with labeled and unlabeled nucleotides. CAGE is another sensitive method but similar to GRO-cap or PRO-cap it does not map transcripts in their full length but only their 5' ends (Shiraki et al, 2003).

Dynamic transcriptome analysis (DTA; Miller et al, 2011), also called 4sU-Seq is the method of choice when it comes to mapping transient RNAs in their full-length and additionally learning about synthesis and decay rates. To be able to transfer this method from yeast to human cells, one has to take into consideration that human genes are much longer than yeast genes. The average gene in yeast is 1.6 kilobase pairs (kb) long and would require 50 to 100

sec to be transcribed with a Pol II synthesis rate of 2 to 4 kb/min (Ardehali & Lis, 2009; Darzacq et al, 2007; Singh & Padgett, 2009). In 6 min labeling time, as employed in the DTA protocol, an average gene in yeast can be labeled between 4 to 7 times. The majority of isolated labeled RNA will therefore be completely labeled.

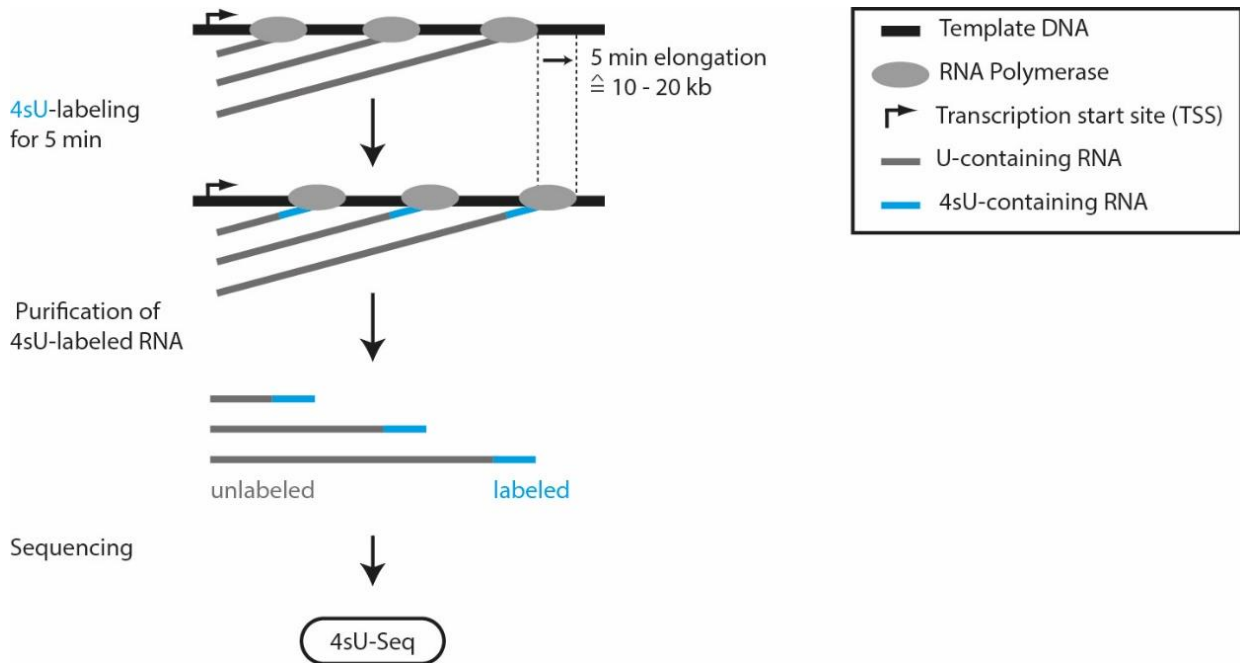


Figure 2: Schematic of 4sU-seq protocol when applied to human cells. If transferred as such from yeast to human, a large portion of sequenced RNAs will in fact be unlabeled RNAs and not labeled RNAs.

In human cells, a protein-coding gene has a mean length of 67 kb. Assuming an average elongation of 2 to 4 kb/min (Ardehali & Lis, 2009; Darzacq et al, 2007; Singh & Padgett, 2009), this implies that in a short labeling time of 5 min Pol II can transcribe a maximum of 20 kb. An average gene will therefore be labeled at an extent of ~ 30% during the labeling time of 5 min. Furthermore, in a steady-state experiment, transcription would be ongoing before the start of the experiment and addition of 4sU (Figure 2). Additionally, Pols are distributed over the gene body in different positions before the 4sU is added - some just initiating, some elongating and some could be already at the 3' end of the gene. Therefore, affinity purified labeled RNA will consist on average of up to 70% of unlabeled RNA. This has two disadvantages: first, the synthesis rate will be overestimated and second, sensitivity for smaller and less abundant transcripts (ncRNAs mostly) is greatly reduced, since very large

unlabeled protein-coding RNA will take up most of the sequencing depth. Taken together, a method is required to be able to study ncRNAs *in vivo*.

1.5 Aims of this thesis

Since the discovery that the majority of the human genome is transcribed (Wilhelm et al, 2008), the question arose if all transcripts are functional or if ncRNAs are a mere by-product of the transcription of protein-coding genes (Palazzo & Lee, 2015). Furthermore, since the discovery of eRNAs, it is still an open question in the field if eRNAs are functional or not, and how their transcription is coupled to their target promoter (Li et al, 2016). Most of the functions, occurrence, abundance and kinetics of ncRNAs in cells remain elusive due to the lack of an adequate experimental protocol. Most of the methods used so far are either *in vitro*, not genome-wide, and/or under perturbed cellular conditions such as deletion or depletion of cellular factors. Furthermore, these methods lack sensitivity and/or do not map RNAs in their full-length but are rather biased towards their 5' or 3' end (Arner et al, 2015; Core et al, 2014; Kwak et al, 2013; Preker et al, 2008). The aim of this thesis was to develop a superior method that would answer the above listed questions.

In the first part of this thesis, we aimed to improve and adapt the 4sU-seq protocol to be able to analyze transient RNA with high sensitivity in human cells (Section 3.1; Miller et al, 2011). With this protocol we wanted to map, annotate and further characterized ncRNAs in a steady-state setup. Of special interest were synthesis and degradation rates of special ncRNAs classes, such as eRNA. Furthermore, we wanted to analyze the sequence composition of eRNAs and investigate if it could explain their unstable nature. Next, through the use of an ENCODE cell line and the available datasets we wanted to characterize differences between promoters and enhancers and their binding factors. Additionally, we wanted to learn more about termination of Pol II. Termination sites can be kb away from the pA sites and we wanted to investigate if we made the same observation. Furthermore, we wanted to investigate the exact nucleotide composition of termination sites to see if our improved sensitivity could let us find an exact motif that would enhance termination. Last, if such a motif were to be found, we wondered if it would depend on a functional pA site.

In the second part of the thesis, we aimed to learn more about eRNA expression kinetics (Section 3.2). We performed a high-resolution time-series of the first 15 min of T cell activation and wanted to analyze expression changes of mRNAs and ncRNAs. To this end, we wanted to obtain a Jurkat cell specific annotation of all transcripts expressed before or during the 15 first minutes of the T cell activation. Next, we wanted to perform differential expression analysis of all transcripts and uncover very rapid responsive transcripts that were

not described in previous studies. Furthermore, to be able to characterize promoter and enhancer pairs, we wanted to employ a stringent pairing method that would minimize the number of false positive pairs. With those pairs we wanted to analyze characteristics of enhancer promoter pairs, such as expression correlation, distance and in particular activation or shutdown kinetics. The question if eRNA expression had to precede mRNA expression for correct transcript activation was of particular interest. Last, we wanted to analyze the CTD modification status of Pol II on our newly annotated ncRNAs and see if we find differences between eRNAs and mRNAs.

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial strain

Table 1: Bacterial strain used in this work (used in section 2.2.5)

Strain	Description	Source
XL-1 Blue	Rec1A; endA1; gyrA96; thi-1; hsdR17; supE44; relA1; lac[F ⁺ proAB lacIqZDM15Tn10(Tetr)]	Stratagene (La Jolla, CA)

2.1.2 Human cell lines

Table 2: Human cell lines used in this work (used in sections 2.2.1, 2.2.3, 2.2.4 and 2.2.5)

Cell line	Cell type	Source
K562	Chronic myeloid leukemia in blast crisis	DSMZ (Braunschweig)
Jurkat	T cell leukemia	Conzelmann group, Gene Center, LMU Munich

2.1.3 Media and supplements

Table 3: Media used in this work (used in sections 2.2.1, 2.2.3, 2.2.4, and 2.2.5)

Name	Description	Application
LB	1% (w/v) tryptone; 0.5% (w/v) yeast extract; 0.5% (w/v) NaCl; (+1.5% (w/v) agar for solid media plates)	<i>E.coli</i> culture
RPMI-1640	RPMI-1640; 10% FBS; 1% Pen-Strep	Human cell culture

Table 4: Additives used in this work (used in sections 2.2.1, 2.2.2, 2.2.3, 2.2.4 and 2.2.5)

Abbreviation	Name	Description	Working concentration
Amp	Ampicillin	Antibiotic	100 µg/ml

Abbreviation	Name	Description	Working concentration
FBS	Fetal bovine serum	Serum supplement	10 %
Pen-Strep	Penicillin-streptomycin	Antibiotic	100 µg/ml
PMA	Phorbol 12-myristate 13-acetate	Activates Protein Kinase C for T cell activation	50 mM
-	Ionomycin	Calcium ionophore for T cell activation	1 µM
4sU	4-thiouridine	Metabolic labeling of nascent RNAs	500 µM

2.1.4 Spike-ins

Table 5: Spike-ins used in this work (used in 2.2.2.2 and 2.2.3)

Spike name	ERCC ID	GenBank ID	Length [nt]	Uridine amount	GC content [%]	comment
2	ERCC-00043	DQ516787	1023	303	33	4sU
12	ERCC-00170	DQ516773	1023	316	34	no 4sU
4	ERCC-00136	EF011063	1033	268	42	4sU
5	ERCC-00145	DQ875386	1042	266	44	no 4sU
8	ERCC-00092	DQ459425	1124	296	50	4sU
9	ERCC-00002	DQ459430	1061	266	51	no 4sU

2.1.5 Primers

Table 6: Spike-ins PCR primers used in this work (used in 2.2.2.2)

Name	Sequence (5'-3')	Annealing temperature [°C]
Spike 2 fwd	TAATACGACTCACTATAGGGTGCTTTAACAAGAGGAAATTGTGT	53
Spike 2 rev	CCATCTTGTTTATAAAATCCTAATTACTC	53
Spike 12 fwd	TAATACGACTCACTATAGGGGGCACAAGTTGCTGAAGTTGC	55
Spike 12 rev	TCTGCTGTAATCTCAGCTCC	55
Spike 4 fwd	TAATACGACTCACTATAGGGTTTCGACGTTTTGAAGGAGGG	53

Spike 4 rev	GTACCCGGGAAAATCCTAGTTC	53
Spike 5 fwd	TAATACGACTCACTATAGGGACTGTCCTTTCATCCATAAGCGG	55
Spike 5 rev	CGCACGCCGAATGATGAAACG	55
Spike 8 fwd	TAATACGACTCACTATAGGGGATGTCCTTGGACGGGGT	55
Spike 8 rev	GCTTTCGGAGCAAATCGCG	55
Spike 9 fwd	TAATACGACTCACTATAGGGGATGTCCTTGGACGGGGT	55
Spike 9 rev	GGGTAAAACGCAAGCACCG	55

Table 7: qPCR primers for labeling protocol efficiency used in this work (used in 2.2.3.5)

Name	Sequence (5'-3')
Spike-in 4 fwd	CCGAGTTCGCCTTACTGCTC
Spike-in 4 rev	AATCGATCGGAATCACGCCG
Spike-in 12 fwd	AGACTGGCATTCCCGTGATA
Spike-in 12 rev	GCTAAAACCCCTGCCTGCAA

Table 8: qPCR primers used during this thesis for ChIP-qPCR experiments (used in 2.2.4.7)

Name	Sequence (5'-3') fwd primer	Sequence (5'-3') rev primer	Binding site
FOS 1	TCCTCCACCCTCCAAGT	TGAGGGCTGGTTCGGTC	5' UTR
FOS 2	TTCCACGCTTTGCACTGAA	GGACTTAAGCCTCGGCTC	promoter
FOS 3	GAGCCCGTGACGTTTACT	GGCTCAGTCTTGCTTCTCA	TSS
FOS 4	GTCAACGCGCAGGTAAGG	GGGAGCCCCCTACTCATCTA	intron 1
FOS 5	CCATGACAGGAGGCCGA	ACATTCCCAGGAAGAGTACG	exon 2/intron 2
FOS 6	CCGAGGCTGACTCCTTCC	GTCAGAGGAAGGCTCATTGC	exon 4
FOS 7	GCACAAATAATGGCTGATCGT	TTCAGAGTCATGTTGACTTCTCC	3' UTR
GAPDH 1	CAGCCAGACGAGGACACA	CCTTTCTGGGATTGCCTTTC	5' UTR
GAPDH 2	ATCCAAGCGTGTAAGGGT	GGAAGGGACTGAGATTGGC	promoter
GAPDH 3	CCCCGGTTTCTATAAATTGAGC	GGCTGACTGTCGAACAGGA	TSS
GAPDH 4	CAGAAACAGGAGGTCCCTAC	GCGCGAAAGGAAAGAAAGC	intron 2
GAPDH 5	ATAGGCGAGATCCCTCAA	TGAAGACGCCAGTGGAC	exon 5
GAPDH 6	GCTGGCACCACTACTTCAGA	TGTTTGGCCAACAGCAGATA	3' UTR

Name	Sequence (5'-3') fwd primer	Sequence (5'-3') rev primer	Binding site
GAPDH 7	AGATGTGTCAGGGTGACTTAT	TAGGTCCCAGCTACACGC	3' UTR
ACT 1	GGCTGGGCGTGACTGTTA	GGGCAGGGCACCTTTTAC	promoter
ACT 2	CCGAAAGTTGCCTTTTATGG	CAAAGGCGAGGCTCTGTG	TSS
ACT 3	CGGGGTCTTTGTCTGAGC	CAGTTAGCGCCCAAAGGAC	exon 3
ACT 4	CGCCCTTTCTCACTGGTTC	TCCAAAGGAGACTCAGGTCAG	intron 4
ACT 5	AAGTCCCTTGCCATCCTAAAA	ATGCTATCACCTCCCCTGTG	exon 6
ACT 6	CCTGCAGACAGCTCTGGTTA	TCTCCAGTCACAAGGCAGAA	3' UTR
IL-2 1	GGCCATAGGGCTGATAAAGA	AAATCTGAATGAGCTCGGACAT	5' UTR
IL-2 2	GGTTTAAAGAAATTCCAAAGAGTCA	TTCAAAGACTTTACCTGTCTGAAAAA	promoter
IL-2 3	ACCTCAACTCCTGCCACAAT	GCAAGACTTAGTGCAATGCAAG	TSS
IL-2 4	GGATTACAGATGATTTGAATGGA	TCCTGGTGAGTTGGGATTTC	exon 1/intron 1
IL-2 5	AGGCCACAGAACTGAAACATC	AAGTGAAAGTTTTTGCTTTGAGC	exon 3
IL-2 6	TCAGAGGAAAAGCGATCAAGT	TCCCTAAATTGGGTGCAAAT	intron 3
IL-2 7	GCCCTTTGGCTTTCCATT	GTCTCTCATAGTTCCTACCAGCA	3' UTR
IFNg 1	AGTCATCCAATGTGCCAAAAT	GGTAATCCTCATAAAGTGCTAGGAA	5' UTR
IFNg 2	CCTCAGGAGACTTCAATTAGGTATAAA	TCCAAAGGACTTAACTGATCTTTCTC	TSS
IFNg 3	TCCAAGGAGAGTGACAGAAAAAT	TTTGGATGCTCTGGTCATCTT	exon 3
IFNg 4	TGCTCAGCTTTCCTATTGCTG	TTTCTGGGGGCTTACATGAG	intron 3
IFNg 5	AGTGAGGAGATGCAAGTAGTTCAA	AAGAATTGCAACAACCTTTCCAG	intron 3
IFNg 6	TGCTGTGTTGGACTTTTTCTAAGT	AGACTCCCCTCCCTACTAATTCA	3' UTR
IFNg 7	GGGGTGGAGTAGTGGGTGTA	CACACTACTTGTAATGGAATAACTTGG	3' UTR
MYC 1	AAGACGCTTTGCAGCAAAATC	AGGCCTTTGCCGCAAAC	5' UTR
MYC 2	GTAGTTAATTCATGCGGCTCTCTTACT	GGGCAGCCGAGCACTCTA	promoter
MYC 3	TTTATAATGCGAGGGTCTGGA	AGAAGCCCTGCCCTTCTC	TSS
MYC 4	GCCGCATCCACGAAACTTT	TCCTTGCTCGGGTGTTGTAAG	exon 1
MYC 5	TGCCCCCTCAACGTTAGCTTC	GGCTGCACCGAGTCGTAGTC	exon 2
MYC 6	ACACAATGTTTCTCTGTAAATATTGCCA	ACTAGGATTGAAATTCTGTGTAAGTCTG	exon 3
MYC 7	GATGCTTCTGAGACTATGATAACA	GCCTTCTGCCATTCTTCTAACT	3' UTR
IL3R α 1	ACCACCCCTGACCACTAAGA	GCGGAGGAGGTTACAGTGAA	promoter
IL3R α 2	CAAGGAAAGTCAGGTTTCATGGT	GCTGTGAGAGACCCCTTCT	TSS
IL3R α 3	ATGGTCTCTCTTTGGCTCA	CTTCTTCGTTTGCAGGAGA	exon 2
U2 1	GGCCTCGTCTTTCTCCAGGA	GGGTCTGTGGAAGAGACTGTCTG	promoter

Name	Sequence (5'-3') fwd primer	Sequence (5'-3') rev primer	Binding site
U2 2	TGGATGAGAGTGGGACGGTG	AGCCAAAAGGCCGAGAAGC	TSS
U2 3	GGGAGGTCCTTAGGATCTCAGC	GAGAGACAGGAGAAAGAAGAACTCCG	3'UTR
H4 1	TTCTTGTCCTTTCTTGTATTTCCACAGT	TACCACGTTATGAGGCTTTAAAAAATTGCT	promoter
H4 2 fwd	GACTCCTCTTGCTCGTCATGTCTG	CGCCTTTGCCAAGACCCCT	TSS
H4 3	GGTTGAGCGTCCCTTTCTATCAACA	TGGGCAAACAAGCATCACGG	exon 1
Neg 1	CGGTAAGAGACCCCTCACAA	CACACTGACTCTCAACAGACTCC	intergenic
Neg 2	CTGATGTCCAGGAGGAGAAAGG	AGCCCGACAATGTCAAGGACTG	intergenic

Table 9: Cloning primers used in this work (used in 2.2.5.2)

Name	Sequence (5'-3')	5' phospho-rylation
Insert 1 control fwd	AGTATCGAAGTCAGCAACTGgcggtgggctctatggctt	yes
Insert 2 control fwd	AGTATCGAAGTCAGCAACTGcccgtcctttcgctttcttcc	yes
Insert 3 control fwd	AGTATCGAAGTCAGCAACTGcttgattagggtgatggttcacgtagtg	yes
Insert 4 control fwd	AGTATCGAAGTCAGCAACTGttataagggattttggggatttcggc	yes
Insert 1 C ₃ AN ₈ T ₄ fwd	AGTACCCAAGTCAGCATTTTgcggtgggctctatggctt	yes
Insert 2 C ₃ AN ₈ T ₄ fwd	AGTACCCAAGTCAGCATTTTcccgtcctttcgctttcttcc	yes
Insert 3 C ₃ AN ₈ T ₄ fwd	AGTACCCAAGTCAGCATTTTcttgattagggtgatggttcacgtagtg	yes
Insert 4 C ₃ AN ₈ T ₄ fwd	AGTACCCAAGTCAGCATTTTttataagggattttggggatttcggc	yes
Insert 1 C ₇ AN ₈ T ₄ fwd	CCCCCCCCAAGTCAGCATTTTgcggtgggctctatggctt	yes
Insert 2 C ₇ AN ₈ T ₄ fwd	CCCCCCCCAAGTCAGCATTTTcccgtcctttcgctttcttcc	yes
Insert 3 C ₇ AN ₈ T ₄ fwd	CCCCCCCCAAGTCAGCATTTTcttgattagggtgatggttcacgtagtg	yes
Insert 4 C ₇ AN ₈ T ₄ fwd	CCCCCCCCAAGTCAGCATTTTttataagggattttggggatttcggc	yes
Insert 1 rev	tcccagcatgcctgctattg	no
Insert 2 rev	cgctagggcgctggcaag	no
Insert 3 rev	tcgaggtgccgtaaagcactaaatc	no
Insert 4 rev	atcaaaagaatagaccgagatagggttgag	no
Remove pA fwd	ttcttactgtcatccaagtaagatgctt	yes
Remove pA rev	gcggccgcttagtcacca	no

Table 10: qPCR and colony PCR primers used for the termination experiment (used in 2.2.5.2 and 2.2.5.3)

Name	Sequence (5'-3')
Term qPCR 1 fwd (Transfection efficiency readout fwd)	CCTCAAAGGGCTTGCCAACG
Term qPCR 1 rev (Transfection efficiency readout rev)	CCTTGATCTTGTCCACCTGGC
Term qPCR 2 fwd	CTGAGGCGGAAAGAACCAGC
Term qPCR 2 rev	CTGGCAAGTGTAGCGGTCAC
Term qPCR 3 fwd	CCTTTCGCTTTCTTCCCTTCCTTTC
Term qPCR 3 rev	CGTAAAGCACTAAATCGGAACCCCT
Term qPCR 4 fwd	ACTTGATTAGGGTGATGGTTCACG
Term qPCR 4 rev	GGTTGAGTGTTGTTCCAGTTTGGA
Term qPCR 5 fwd (Termination read-through readout fwd)	GTGGAATGTGTGTCAGTTAGGGTG
Term qPCR 5 rev (Termination read-through readout rev)	GACTTTCACACCTGGTTGCT

2.1.6 Thermal cycler programs

Table 11: Thermal cycler programs used in this work (for section see details of table)

Program	Name	Used in section
25°C, 10 min; 50°C, 30 min; 85°C, 5 min	Reverse-transcription	2.2.3.5, 2.2.5.3
95°C, 2 min ; (95°C, 20 sec; 53-55°C, 30 sec; 72°C, 70 sec) x30; 72°C, 10 min; 10°C, ∞	Spike-in PCR	2.2.2.2
95°C, 2 min ; (95°C, 5 sec; 64°C, 10 sec; 72°C, 15 sec) x40; 72°C, 5 min; 10°C, ∞	qPCR	2.2.3.5, 2.2.4.7, 2.2.5.3
98°C, 2 min ; (98°C, 45 sec; 60°C, 30 sec; 72°C, 6 min) x29; 72°C, 10 min; 10°C, ∞	Cloning PCR	2.2.5.2
95°C, 2 min ; (95°C, 5 sec; 60°C, 10 sec; 72°C, 7 sec) x40; 72°C, 5 min; 10°C, ∞	Colony PCR	2.2.5.2

2.1.7 Plasmids

Table 12: Plasmids used in this work (generated and used in section 2.2.5)

Name	Description	Source
pCMV-GLuc 2	CMV; GLuc; SV40; Neo ^R ; pMB1 ori; Amp ^R	NEB
pCMV-GLuc 2 4x control	pCMV-GLuc 2 + 4x AGTATCGAAGTCAGCAACTG	This work

Name	Description	Source
	inserted between GLuc pA and SV40	
pCMV-GLuc 2 4x C ₃ AN ₈ T ₄	pCMV-GLuc 2 + 4x AGTACCCAAGTCAGCATTTT inserted between GLuc pA and SV40	This work
pCMV-GLuc 2 4x C ₇ AN ₈ T ₄	pCMV-GLuc 2 + 4x CCCCCCAAGTCAGCATTTT inserted between GLuc pA and SV40	This work
pCMV-GLuc 2 no pA 4x control	pCMV-GLuc 2 4x control + pA of GLuc removed	This work
pCMV-GLuc 2 no pA 4x C ₃ AN ₈ T ₄	pCMV-GLuc 2 4x C ₃ AN ₈ T ₄ + pA of GLuc removed	This work
pCMV-GLuc 2 no pA 4x C ₇ AN ₈ T ₄	pCMV-GLuc 2 4x C ₇ AN ₈ T ₄ + pA of GLuc removed	This work

2.1.8 Buffers and solutions

Table 13: Buffers and solutions used in this work (for section see details of table)

Name	Description	Application/ section
1x TBE	8.9 mM Tris-HCl; 8.9 mM Boric acid; 2 mM EDTA; pH 8.0 at 25°C	Agarose gel/ 2.2.2.3, 2.2.2.4,
6 x Loading dye	1.5 g/L Bromphenol blue; 1.5 g/L Xylene cyanol; 50% (v/v) Glycerol	2.2.4.3, 2.2.5.2
10x biotinylation buffer	100 mM Tris pH 7.4; 10 mM EDTA	
Biotin-HPDP	1mg/ml in dimethylformamide	TT- seq/ 2.2.3.4
Washing buffer	100 mM Tris pH 7.5; 10 mM EDTA; 1 M NaCl; 0.1% Tween20	
Elution buffer (DTT)	100 mM Dithiothreitol in H ₂ O	
Protease-inhibitor mix	1 mM Leupetin; 2 mM Pepstatin A; 100 mM Phenylmethylsulfonylfluoride; 280 mM Benzamidine	
Phosphatase-inhibitor mix	1 mM NaN ₃ ; 1 mM NaF; 0.4 mM Na ₃ VO ₄	ChIP-seq/ 2.2.4
1 x PBS	2 mM KH ₂ PO ₄ ; 4 mM Na ₂ HPO ₄ ; 140 mM NaCl; 3 mM KCl; pH 7.4 at 25°C	
Farnham Lysis buffer	5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40	
FA lysis buffer 1% SDS	50 mM HEPESKOH, pH 7.5 at 4°C; 150 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (v/v) Na deoxycholate; 1% (v/v) SDS; PI	ChIP-seq/ 2.2.4.2
PBS/Tween	0.02% Tween20 in PBS	ChIP-seq/ 2.2.4.4, 2.2.4.5
Co-IP buffer	0.5% NP-40, 150 mM NaCl, 50 mM Tris pH 7.4, 2 mM EDTA	ChIP-seq/ 2.2.4.5

Name	Description	Application/ section
LiCl wash buffer	100 mM TrisHCl, pH 8.0 at 4°C; 500 mM LiCl; 1% (v/v) NP-40; 1% (v/v) Na deoxycholate	ChIP-seq/ 2.2.4.6
TE Buffer	10mM TrisHCl, pH 7.4 at 4°C; 1mM EDTA	
ChIP elution Buffer	0.1 M NaHCO ₃ ; 1% (v/v) SDS	

2.1.9 Antibodies

Table 14: Antibodies used in this work for ChIP experiments (used in section 2.2.4.4)

Antibody	Amount in ChIP	Source
N-20	5 µl	Santa-Cruz
3D12 (Tyr1P)	50 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
3E8 (Ser5P)	20 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
3E10 (Ser2P)	25 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick
4E12 (Ser7P)	50 µl (supernatant cell culture)	Elisabeth Kremmer/Dirk Eick

2.2 Experimental methods

In this section, only methods developed and performed as part of this thesis are described. Methods performed by other co-authors regarding Section 3.1 and 3.2 are listed in the supplementary materials of the papers (Michel et al, 2016; Schwalb et al, 2016).

2.2.1 Human cell culture

K562 cells were acquired from DSMZ (Braunschweig, Germany; Table 2). Jurkat cells were kindly provided by the Conzelmann lab, Gene Center, LMU Munich (Table 2). Cells were grown in RPMI-1640 medium (Gibco, Carlsbad, CA, USA) supplemented with 10% heat-inactivated FBS (Gibco) and 1% Penicillin/Streptomycin (100x, PAA now GE Healthcare, Chalfont St. Giles, UK) at 37°C under 5% CO₂ (Table 3, Table 4). Cells were splitted 1:4 with fresh media every second day.

2.2.2 Generation of spike-ins

2.2.2.1 Background

Spike-ins are crucial for normalization when comparing different samples within an experiment. They are synthetic probes that do not match any sequence in the human genome in order to clearly differentiate them from sequences from your biological experiment. Spike-ins are added at the beginning of an experiment to every sample in the same amount. They serve as an internal standard and can correct for global changes in RNA amount between different samples. The idea to use spike-ins is based on a similar approach developed in the Cramer lab for *S. cerevisiae* (Sun et al, 2012).

Spike-ins used in a 4sU-labeling experiment must be labeled with 4sU so that they are not lost throughout the experiment. Spike-ins were therefore based on commercially available unlabeled RNA spike-ins that were converted to DNA. These converted spike-ins were used as input in an *in vitro* transcription experiment using 4sUTP as a uridine source, resulting in labeled RNA spike-ins. To normalize for labeled and unlabeled RNA, half of the spike-ins were not labeled with 4sUTP.

2.2.2.2 Spike-ins PCR

Six spike-ins from the ERCC RNA spike-in mix (Life Technologies, Carlsbad, CA, USA) were selected by Katja Frühauf based on the following criteria: a length of approximately 1 kb and similar uridine amount but different GC content (40 to 60%; Table 5, Schwalb et al, 2016). Spike-ins were extracted from the ERCC spike-in mix and converted to DNA by PCR amplification with the forward primer containing a T7 promoter sequence (Table 6). The PCR reaction contained 50 ng RNA, 5 µl 10 x PCR buffer (NEB, Ipswich, Massachusetts, USA), 2 µl MgSO₄ (NEB), 5 µl dNTPs (2mM each, NEB), 3 µl forward primer (5 pmol/µl), 3 µl reverse primer (5 pmol/µl), 1 µl Phusion polymerase (NEB) and ddH₂O in a total volume of 50 µl. See Table 11 “spike-in PCR” for thermal cycler conditions. Specific annealing temperature for each primer can be found in Table 6. PCR reactions were purified with QIAquick PCR purification columns (Qiagen, Venlo, Netherlands) and eluted in 30 µl ddH₂O.

2.2.2.3 *In vitro* transcription

After assessing quality, concentration and homogeneity of the PCR products (expected size: 1 kb) with a 4% agarose gel (Figure 3, Table 13) and a Nanodrop 2000 (Thermo Fisher Scientific, Boston, MA, USA), each DNA spike-in was subjected to *in vitro* transcription with the Megascript T7 Transcription Kit (Life Technologies). The transcription reaction was performed with either 1:10 4sUTP:UTP ratio for spike-ins 2, 4 and 8 or only UTP for spike-ins 12, 5 and 9 (Table 5); resulting in labeled and non-labeled spike-ins, respectively. The *in vitro* transcription reaction was mixed following the manufacturer's instructions starting with 8 µl/500 ng of DNA and incubated for 4h at 37°C. The reactions were diluted to 40 µl with ddH₂O. To digest input DNA 2 µl Turbo DNase (Thermo Fischer Scientific) and 4 µl 10x Turbo DNase Buffer were added to each reaction and incubated for 15 min at 37°C. No stop solution was added to the reactions.

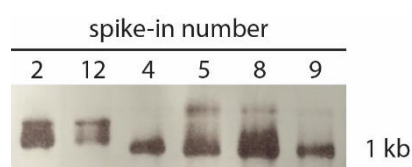


Figure 3: Quality control of RNA spike-ins after *in vitro* transcription. 1 µg of spike-ins were loaded on a 4% agarose gel and run at 100 V for 15 min.

2.2.2.4 Quality assessment

All RNA spike-ins were purified with AMPure XP beads with a 1.8x ratio (Beckman-Coulter, Brea, CA, USA), eluted in 400 µl ddH₂O and quantified with Nanodrop 2000, agarose gel and Qubit 3.0 Fluorometer (Life Technologies). A biotinylation-streptavidin purification (see 2.2.3.4 for details) was performed with each spike-in to check for correct 4sU-labeling or non-labeling. The spike-ins were then mixed together as a stock for all future experiments in equal amount (1 ng/µl for each spike) to a final concentration of 6 ng/µl and stored at -80 °C.

2.2.3 TT-seq

2.2.3.1 Cell culture

K562 or Jurkat cells were splitted 2 days prior to TT-seq experiment and kept at a maximum cell density of 1×10^6 cells/ml to ensure optimal growth conditions. Experiments were performed with cells below passage 10 in T75 flasks (Greiner Bio-one, Kremsmünster, Austria).

2.2.3.2 RNA purification

50 million cells were centrifuged for 5 min at 2,000 rpm and resuspended in 50 ml pre-warmed media to obtain equal labeling conditions for each sample and experiment. Labeling was performed in media for 5 min with 500 μ M of 4sU (50 mM stock solution in water; Sigma-Aldrich, St. Louis, MO, USA; Table 4) after which cells were harvested through centrifugation for 2 min at 3,000 rpm. 20 μ l of RNA spike-in mix (see section 2.2.2) was added to each cell pellet together with 5 ml Qiazol (Qiagen). RNA extraction was performed following the Qiazol protocol's instructions. The extracted RNA was resuspended in 500 μ l ddH₂O and is referred to as "Total RNA". Total RNA concentration and purity was measured with Nanodrop 2000.

2.2.3.3 RNA fragmentation

Trials of fragmentation were performed in a time series on 400 μ g of Total RNA for 5 min with high settings (Figure 4A) and on Poly A + RNA (purified with Ambion Poly(A) Purist MAG Kit) for 7 min on high settings or 4 min on low settings (Figure 4B) on a BioRuptor Next Gen (Diagenode, Seraing, Belgium) using 1.5 ml Bioruptor Plus TPX microtubes (Diagenode). The time series were analyzed on an Experion (Bio-Rad, Hercules, CA, USA; Figure 4). Fragmented RNA was supposed to be in a range below 10 kb and above 200 bp. Base on that criterion we decided to apply 1 min at high settings for all subsequent fragmentations.

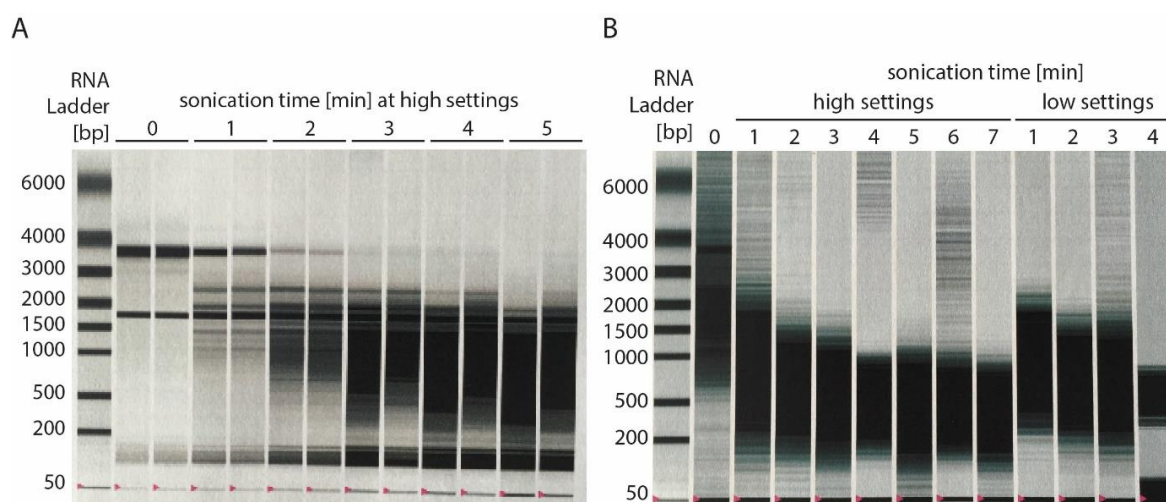


Figure 4: Electrophoresis analysis of fragmentation time series on a BioRuptor NextGen. (A) Fragmentation on Total RNA at high settings for different time. (B) Fragmentation time series on Poly A+ RNA using different settings.

In K562 or Jurkat cells, 300 µg of Total Fragmented RNA are needed per 5 min 4sU-labeling reaction to obtain a final yield of approximately 300 ng of labeled fragmented RNA. 300 µg Total RNA was fragmented at 240 ng/µl in a maximum of 400 µl per 1.5 ml Bioruptor Plus TPX microtubes on a BioRuptor Next Gen at high power for one cycle of 30''/30'' ON/OFF. The resulting RNA is referred to as “Total fragmented RNA“. Fragmentation efficiency was assessed on a Tapestation (Agilent Technologies, Santa Clara, CA, USA) by comparing Total RNA and Total fragmented RNA samples using a RNA ScreenTape (Agilent Technologies; Figure 5). Total fragmented RNA samples were then subjected to labeled RNA purification (Dolken et al, 2008).

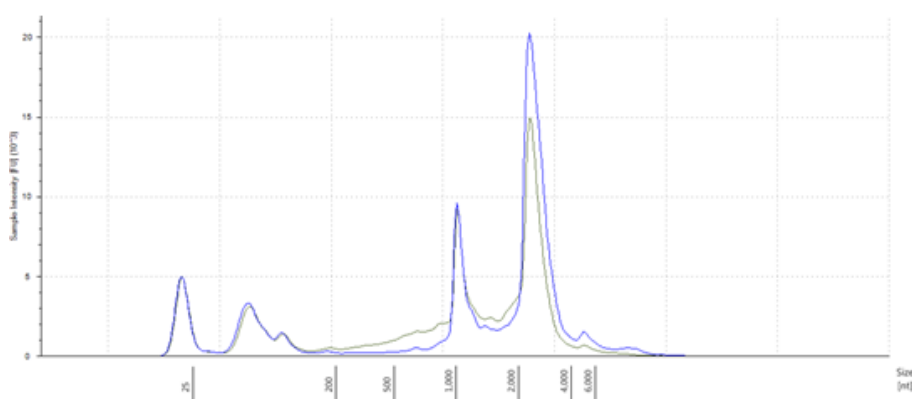


Figure 5: Comparison between Total RNA and Total fragmented RNA. Overlaid electropherograms acquired on a Tapestation of 100 ng Total RNA (blue) and 100 ng Total fragmented RNA (grey) after 1 min of fragmentation on a BioRuptor.

2.2.3.4 Labeled RNA purification

Briefly, 300 µg Total fragmented RNA was first heated to 60°C for 10 min and put on ice for 2 min. The RNA was then biotinylated in 15 ml RNase-free falcons (Sarstedt, Nümbrecht, Germany) in a biotinylation reaction containing 300 µl 10x Biotinylation Buffer (1 µl per 1 µg RNA) and 600 µl Biotin-HPDP (2 µl per 1 µg RNA; Thermo Fischer Scientific, Table 13). The biotinylation reaction was brought to a final volume of 2.1 ml with ddH₂O (7 µl per 1 µg RNA), kept from light and incubated on a rotating wheel for 1.5 hours at room temperature (RT). Unbound biotin was removed by chloroform/isoamylalcohol extraction in 15 ml phase lock gel tubes (5 PRIME, Hilden, Germany). The biotinylated RNA was then precipitated with 1 volume isopropanol and 1/10th volume 5 M NaCl. The biotinylated RNA was washed twice with 70% ethanol and resuspended in 200 µl RNase-free water. The biotinylated RNA was heated to 65°C for 10 min and put on ice for 5 min. 100 µl Streptavidin µMACS

paramagnetic beads (Miltenyi Biotec, Bergisch Gladbach, Germany) were added to the biotinylated RNA and incubated on a shaker at 800 rpm for 15 min at 24°C. The μ MACS column was equilibrated with 900 μ l washing buffer (Table 13) before adding the biotinylated RNA/streptavidin mixture to the column. The flow-through was reloaded to maximize labeled RNA recovery. The column was washed 3 times with 900 μ l 65°C warm washing buffer and 3 times with 900 μ l RT washing buffer. The labeled fragmented RNA was eluted with two times 100 μ l 100 mM DTT (Table 13) with a 5 min incubation time in between the two elutions. The labeled fragmented RNA was purified using 1.8x AMPure XP RNase-free beads and resuspended in 30 μ l RNase-free water. RNA concentration was measured with Qubit (Thermo Fischer Scientific) using the Qubit RNA HS Assay Kit (Thermo Fischer Scientific). Quality can be assessed with a Tapestation using a high-sensitivity RNA ScreenTape (Agilent Technologies, Figure 6).

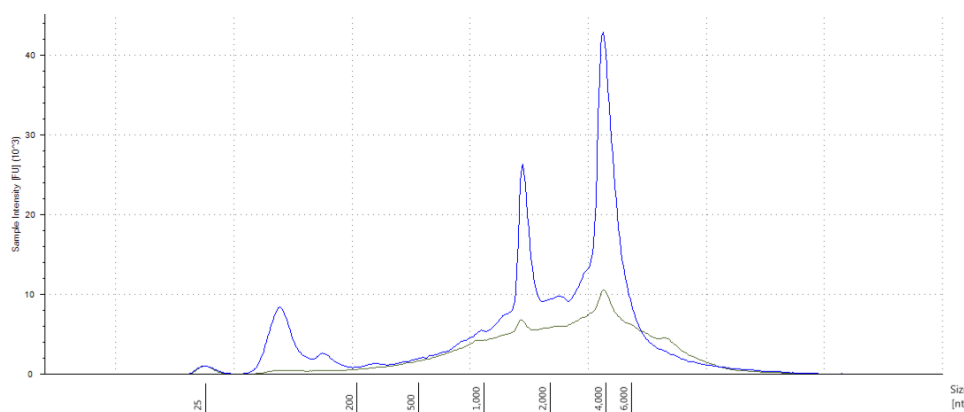


Figure 6: Comparison between Total fragmented and Labeled fragmented RNA. Overlaid electropherograms acquired on a Tapestation of 10 ng Total fragmented RNA (blue) and 10 ng Labeled fragmented RNA (grey) after labeling purification.

2.2.3.5 qPCR against spike-ins to assess labeled RNA purification efficiency

Successful labeled RNA purification was tested through qPCR with primers against a labeled and unlabeled spike-in. The ratio of labeled/unlabeled spike-ins should be 1 in Total RNA or Total fragmented RNA samples and greater than 1 in Labeled fragmented RNA samples. Different starting concentration between Labeled and Total RNA is not a problem, since one does not compare Labeled and Total RNA ratio to one another, but different spike-ins to one another within the same sample. Labeled fragmented, Total fragmented and Total RNA were reverse-transcribed to cDNA in a reaction containing 1 μ l RNA, 1 μ l 50 μ M random hexamers (Thermo Fischer Scientific), 1 μ l 10 mM dNTPs (NEB), 4 μ l 5x buffer (Thermo Fischer Scientific) and 1 μ l Maxima Reverse-transcriptase (200 U/ μ l, Thermo Fischer

Scientific). The reaction was brought to 20 µl with ddH₂O and incubated in a thermal cycler (see Table 11 “Reverse-transcription” for parameters). A qPCR reaction was mixed with 1 µl cDNA, 10 µl 2x SensiFAST SYBR mix (Bioline, London, UK), 0.8 µl each 10 µM forward and reverse primers for spike-ins 4 (labeled; Table 7) or 12 (unlabeled; Table 7) and filled to 20 µl with ddH₂O. See Table 11 “qPCR” for the thermal cycler conditions. The Ct-values between spike-in 4 and 12 should be equal in Total RNA samples and greater Ct-value for spike-in 12 than spike-in 4 in Labeled RNA samples (Figure 7). The enrichment of labeled spike-ins over unlabeled spike-ins usually ranges between 7 to 9 Ct-values, meaning a 100- to 500-fold enrichment of labeled RNA. After amount, quality and labeling purification assessment sequencing libraries can be generated.

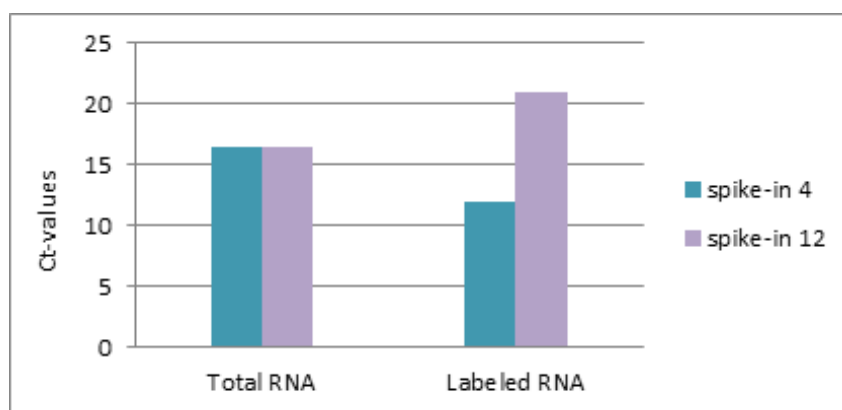


Figure 7: qPCR testing efficiency of labeled RNA purification based on spike-ins enrichment in samples. qPCR was performed on cDNA of Total (left) or Labeled (right) RNA samples with primers testing against spike-in 4 (blue, labeled spike-in) or spike-in 12 (violet, unlabeled spike-in). The Ct-value of spike-in 12 is higher than for spike-in 4 in Labeled RNA sample showing that it is less present in the sample.

2.2.3.6 Generation of sequencing libraries and sequencing

80-150 ng RNA from all samples was treated with 2 µl (2 units) TURBO DNase (Life Technologies) and 0.1 volume 10x TURBO DNase buffer (Life Technologies) for 30 min at 37°C. The reaction was stopped with 0.1 volume DNase inactivation reagent (Life Technologies) and incubated at RT for 5 min. The DNase inactivation reagent was removed by centrifugation at 10,000 rpm for 2 min. Sequencing libraries were prepared with the Ovation Human Blood RNA-seq library kit (NuGEN, Carlos, CA, USA) following the manufacturer’s instructions. If the volume of the 80-150 ng DNase-treated RNA was too large for the sequencing kit requirements, the RNA was centrifuged in a vacuum concentrator until the desired volume was attained. Concentrations during the PCR amplification were measured with a Qubit dsDNA BR assay (Life technologies) from PCR cycle number 13 onward after

each cycle. The PCR reaction was stopped when all samples' concentrations doubled from one cycle to the next. All samples were sequenced on a HiSeq 1500 sequencer (Illumina, San Diego, CA, USA).

2.2.4 ChIP-seq

2.2.4.1 Cell culture and cross-linking

ChIP-seq was carried out as described (Johnson et al, 2007), with some alterations. 3×10^7 Jurkat cells were crosslinked for 7.5 min with 1% methanol-free formaldehyde (Pierce™ 16% Formaldehyde (w/v), Methanol-free, Thermo Fischer Scientific) in 30 ml medium. Glycine was added to a final concentration of 0.125 M and incubated for 5 min at RT on a rotating wheel. Cells were pelleted at 4°C with 1,000 g for 5 min, washed twice with 1 ml 1x PBS (Table 13). Cells were transferred to 1.5 ml Bioruptor Plus TPX microtubes and pelleted at 4°C with 1,000 g for 5 min.

2.2.4.2 Cell lysis and chromatin isolation

From this step forward, proteinase inhibitors were added to all buffers (Table 13). Cells were resuspended in 1 ml Farnham Lysis buffer (Table 13) to lyse the cell membrane and incubated on ice for 5 min. Nuclei were pelleted at 4°C with 2,000 g for 5 min and subsequently resuspended in 1 ml FA 1% SDS (Table 13) to lyse the nucleus' membrane and incubated on ice for 5 min. The chromatin was pelleted at 4°C with 15,000 g for 15 min. The chromatin was carefully resuspended in 300 µl FA 1% SDS (Table 13) so not to generate air bubbles.

2.2.4.3 Sonication and sonication efficiency

The samples were transferred to a pre-cooled BioRuptor Next Gen (4°C) with 3 samples per sonication round at the most. Samples were sonicated with high settings for 40 cycles of 30''/30'' sec ON/OFF. Sonication efficiency was tested by taking 15 µl sonicated chromatin. 100 µl 1x PBS was added to the test samples, together with 1 µl RNase A (10 µg/µl, Thermo Fischer Scientific) and incubated at 37°C for 30 min. 1 µl Proteinase K (NEB) was then added to the test samples and incubated at 37°C for 1 h. Reverse-crosslinking was performed at 65°C at least 4h. Test samples were purified using QIAquick PCR purification columns and eluted in 30 µl ddH₂O. 1 µg test samples were run on a 2% TBE agarose gel (Table 13) and analyzed using a Gel iX20 Imager (Figure 8; Intas Science Imaging Instruments, Göttingen).

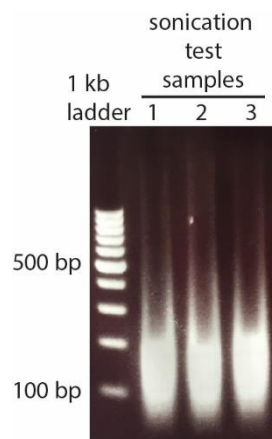


Figure 8: Agarose gel of sonicated chromatin on a BioRuptor NextGen after 40 cycles at high settings 30"/30" ON/OFF.

2.2.4.4 Antibodies-Dynabeads coupling

Per ChIP-seq experiment, 49.5 μ l Dynabeads protein G (Life Technologies) were used. Dynabeads are magnetic beads that can be pelleted with a magnet (Life Technologies). The Dynabeads were washed 3 times with 1 ml PBS/Tween (Table 13) and subsequently resuspended in 1 ml PBS/Tween. The appropriate antibody amount was added to the Dynabeads, depending on the antibody used (for amounts, see Table 14) and incubated on a rotating wheel for 30 min at RT. In case of a mock-IP, no antibody was bound to the Dynabeads. The Dynabeads were then washed three times with 1 ml PBS/Tween and resuspended in 100 μ l PBS/Tween per ChIP-seq sample.

2.2.4.5 Immunoprecipitation

7.5 μ l from the sonicated chromatin was taken as "Input" (1 % of the chromatin), transferred to 0.5 ml DNA LoBind tubes (Eppendorf, Hamburg, Germany) and kept at 4°C for further use. Immunoprecipitation was performed in 13 ml inoculation tubes (Sarstedt) by combining 6 ml Co-IP buffer (Table 13), 100 μ l Antibody-Dynabeads in PBS/Tween and adding 750 μ l sonicated chromatin. Samples were immunoprecipitated overnight at 4°C on a rotating wheel.

2.2.4.6 Reverse-cross-linking and DNA purification

Immunoprecipitated samples were centrifuged for 1 min at 1,000 rpm to gather the Dynabeads at the bottom of the tube and then put on a magnet for 1 min. The supernatant was removed and the Dynabeads were washed 5 times with 1 ml LiCl wash buffer (Table 13) with 3 min incubation time on a rotating wheel at RT in between the washes. The Dynabeads were

then washed once with 1 ml TE Buffer and resuspended in 50 µl IP Elution Buffer (Table 13). To elute the immune-bound chromatin from the Dynabeads, samples were incubated at 70°C for 15 min. The supernatant (“IP” sample) was collected in 0.5 ml DNA LoBind tubes. IP samples were diluted with 45 µL TE buffer; Input samples were complemented with 45 µL IP Elution Buffer and 45 µL TE buffer. 5 µL Proteinase K was added to both the IP and the Input samples, and they were incubated at 37°C for 1 h, then at 65°C overnight or at least for 4h. DNA was purified with MinElute PCR purification columns (Qiagen). 10 µl 3M NaAc was added to all samples before binding them to the columns. DNA was eluted with 3 times 15 µl and concentration was measured with a Qubit HS dsDNA kit (Life technologies). To ensure correct sonication for sequencing, 1 µl Input was analyzed with a DNA 1000 BioAnalyzer kit (Agilent Technologies).

2.2.4.7 ChIP-qPCR

A qPCR was performed to ensure enrichment of the immunoprecipitated DNA over background. Primers were chosen depending on the antibody used, but usually primers binding over the promoter, the TSS, the gene body in exon and introns or near the poly A site were chosen (Table 8). The qPCR reaction was mixed with 1 µl DNA, 10 µl 2x SensiFAST SYBR mix, 0.8 µl each of 10 µM forward and reverse primers (Table 8) and filled to 20 µl with ddH₂O. Cycling conditions can be found in Table 11 “qPCR”. Analysis of the qPCR was performed by calculating the “% Input”: the enrichment of specific DNA sequences in the IP sample over the Input sample:

$$\% \text{ Input} = 100 * 2^{((Ct(\text{Input}) - \log_2(100)) - Ct(\text{IP}))}$$

2.2.4.8 Generation of ChIP-seq libraries and sequencing

The same amount (1-10 ng) Input and IP was taken to generate ChIP-seq libraries using the NEBNext Ultra Kit (NEB) following the manufacturer’s instructions except three modifications. First, The NEB adaptors were diluted 1:20 instead of 1:10 as this reduced the adapter dimers. Second, the DNA was eluted from the AMPure XP beads with the same amount needed in the following step to reduce DNA loss. Last, progress of the PCR reaction was followed by measuring the DNA concentration from cycle 13 onward, similar to the TT-seq library preparation in section 2.2.3.6. to ensure that the PCR is stopped in its exponential

phase to limit PCR duplicates. Samples were sequenced on an Illumina HiSeq 1500 sequencer.

2.2.5 *In vivo* termination experiment

2.2.5.1 Background

Termination motifs (section 3.1) were cloned into a mammalian expression plasmid to test their functionality. It was tested *in vivo* whether the termination motifs have a negative effect on transcription after the pA site of a reporter gene.

The bioinformatically-determined motifs consist of multiples Cs followed by an A and then further downstream multiple Ts (section 3.1). We designed three 20 nucleotides (nt) long inserts: a control insert (Ctrl: AGTA TCGA AGTC AGCA ACTG), an insert with 3 Cs and 4 Ts (C₃AN₈T₄: AGTA CCCA AGTC AGCA TTTT) and an insert with 7 Cs and 4 Ts (C₇AN₈T₄: CCCC CCCA AGTC AGCA TTTT). Since termination windows were found to have multiple termination motifs, it was decided to insert four motifs in the 585 bp region between the poly A site of the *Gaussia* luciferase gene and the SV40 promoter of the neomycin resistance gene of the pCMV GLuc 2 plasmid (NEB). Inserts were not mixed between one another. It was either: four times the Ctrl, 4 times the C₃AN₈T₄ or four times the C₇AN₈T₄ motif (Figure 9). To test if 5' cleavage at the poly A site is a requirement for the termination motifs to act as such, the poly A site of the luciferase gene was deleted in follow-up experiments.

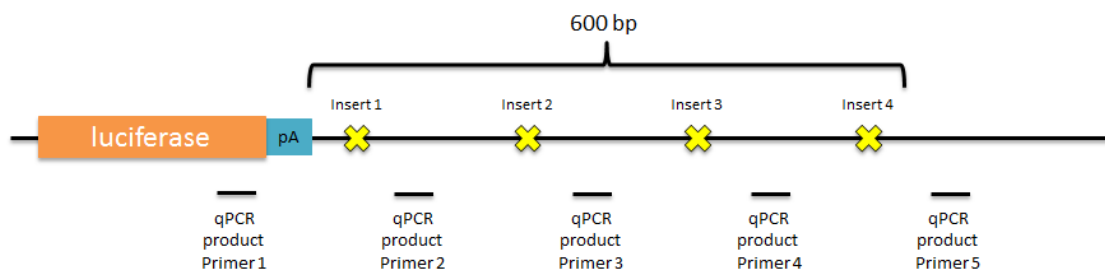


Figure 9: Overview of termination experiment. Yellow crosses represent inserts and are the Ctrl, the C₃AN₈T₄ or the C₇AN₈T₄ motif. In a follow-up experiment, the pA site was deleted.

2.2.5.2 Cloning

Four inserts were inserted into the pCMV-GLuc 2 plasmid (NEB) one after another in four rounds of cloning using the “around the horn PCR” cloning method. The forward primer was

designed to anneal to the plasmid 1 nt downstream of the insertion site, had a 20 nt 5'-overhang consisting of the sequence to be inserted (Table 9) and was phosphorylated on its 5' end to ensure ligation after the PCR. The reverse primer annealed 0 nt upstream of the insertion site (Table 9). The inserts were cloned in the reverse order. First, insert 4 was cloned into pCMV-GLuc 2 plasmid. Second, insert 3 was cloned in the plasmid containing insert 4. Third, insert 2 in the plasmid containing both inserts 4 and 3. Last, insert 1 was cloned in the plasmid containing inserts 4, 3 and 2. All cloning rounds were performed the same: a PCR, overnight DpnI digest, size selection on a 1% agarose gel, overnight ligation of the PCR product into a plasmid and transformation into *E. coli*. The PCR reaction contained 100 ng of plasmid, 2.5 µl 10 µM forward primer (Table 9), 2.5 µl 10 µM reverse primer (Table 9), 1 µl 10 mM dNTPs, 1.5 µl DMSO, 10 µl 5x HF Phusion buffer, 0.5 µl Phusion polymerase and ddH₂O to a final volume of 50 µl. Cycling conditions can be found in Table 11 "Cloning PCR". After PCR, 2 µl of DpnI (NEB) was added to each PCR reaction and incubated overnight at 37°C. 6x Loading dye (Fermentas, Table 13) was added to each reaction and the complete reaction was run on a 1% agarose (Life Technologies) TBE (Table 13) gel at 120 V for 30 min. The correct band (5.8 kb) was excised with a scalpel and purified with a gel purification kit (Qiagen) following the manufacturer's instructions. The linear PCR product was eluted from the Qiagen column in 20 µl ddH₂O. The ligation reaction contained 4 µl linear PCR product, 1 µl 10x T4 DNA Ligase buffer (NEB), 1 µl 10 mM ATP (NEB), 1 µl T4 Ligase (NEB) and 3 µl ddH₂O in a final volume of 10 µl. The ligation reaction was incubated at RT overnight. 4 µl of the ligation reaction was transformed into 50 µl chemically competent XL1-Blue *E. coli* cells (Table 1). Circularized PCR products and XL1-Blue cells were incubated for 30 min on ice, followed by a 30 sec 42°C heat shock in a water bath and 2 min incubation on ice. 900 µl LB medium (Table 3) was added to the cells followed by 1.5 hours incubation at 300 rpm and 37°C. Cells were plated on LB-Amp agar plates (Table 4) and incubated at 37°C overnight. Five colonies were picked the next day and correct insertion (+20 nt) was checked with colony PCR, by directly dipping a colony into a PCR reaction mix containing 5 µl 2x mastermix, 0.5 µl 10 µM forward primer (Table 10), 0.5 µl 10 µM reverse primer (Table 10) and 4 µl ddH₂O. The primers closest to the insertion site were chosen, so that the PCR product was as small as possible to easily differentiate between +/- 20 bp. The colony PCR cycling conditions can be found in Table 11 under "colony PCR". As a negative control, the plasmid taken as a template for the "around the horn PCR" was taken as a template for the colony PCR. The PCR products were run with 6x loading dye on a 2% TBE

agarose gel for 30 min at 100 V. One positive clone was grown overnight in 5 ml LB+Amp at 37°C and 300 rpm, plasmids were extracted the next day using QIAprep spin miniprep Kit (Qiagen), eluted in 50 µl ddH₂O and sent for sequencing at Seqlab (Göttingen) for verification. In the plasmids containing four inserts, the poly A site of the Luciferase gene was deleted following the same cloning strategy with the primers found in Table 9.

2.2.5.3 *In vivo* termination experiment

2 µg plasmid (Table 12) were transfected into 1x10⁶ K652 cells using the SF Cell line 4D-Nucleofector X kit and unit from Lonza (Basel, Switzerland) following the manufacturer's instructions. 500 µl of transfected cells (2.5 x 10⁵ cells) were harvested 4h post transfection by centrifugation for 10 min at 1,900 rpm. 100 µl Qiazol and 20 µl chloroform were added to the cell pellet, incubated at RT for 5 min and centrifuged at 15,000 rpm for 5 min at 4°C. The aqueous supernatant was directly pipetted onto a gDNA Eliminator spin columns (Qiagen) provided in the RNeasy Plus Mini column (Qiagen). This step removed DNA contamination. The RNeasy Plus Mini kit was performed following the manufacturer's instructions. RNA was eluted in 30 µl ddH₂O and concentration was measured with Nanodrop 2000. 500 ng of RNA was reverse-transcribed to cDNA in a reaction containing 1 µl 50 µM random hexamers, 1 µl 10 mM dNTPs, 4 µl 5x buffer and 1 µl Maxima Reverse-transcriptase. The reaction was brought to 20 µl with ddH₂O and incubated in a thermal cycler (see Table 11 "Reverse-transcription" for parameters). A negative control for each sample was pipetted containing all reverse-transcription components except for the Maxima Reverse-transcriptase. This controls for DNA contamination in the samples. A qPCR reaction was mixed with 1 µl cDNA, 10 µl SYBR Select master mix (applied biosystems, Carlsbad, CA, USA), 0.8 µl each 10 µM forward and reverse primers (Table 10) and filled to 20 µl with ddH₂O. See Table 11 "qPCR" for the thermal cycler conditions. The qPCR was performed in technical triplicates. Ct-values of the termination read-through readout (qPCR primer pairs 5, Table 10) were normalized for transfection efficiency using the qPCR readout assessing the luciferase transcript amount (qPCR primer pair 1, Table 10). Additionally, it was controlled for a day-specific effect apparent on the termination read-through readout by setting day-specific control to 0 and normalizing respective values accordingly. Ct values were transformed to absolute numbers by taking it to the power of 2.

3 Results and Discussion

3.1 TT-seq maps the human transcriptome

All results presented in this Section were obtained in collaboration with Björn Schwalb and Benedikt Zacher and are published (Schwalb et al, 2016). For detailed author contributions see page 7. The full article with supplementary materials can be found at <http://science.sciencemag.org/content/352/6290/1225>.

Transcription of eukaryotic genomes produces protein-coding mRNAs and diverse non-coding RNAs (ncRNAs), including enhancer RNAs (eRNAs; Jensen et al, 2013, Andersson et al, 2014). Most ncRNAs are rapidly degraded, difficult to detect, and thus far have not been mappable in their full range. Mapping of transient RNAs is required, however, for analysis of RNA sequence, function, and fate.

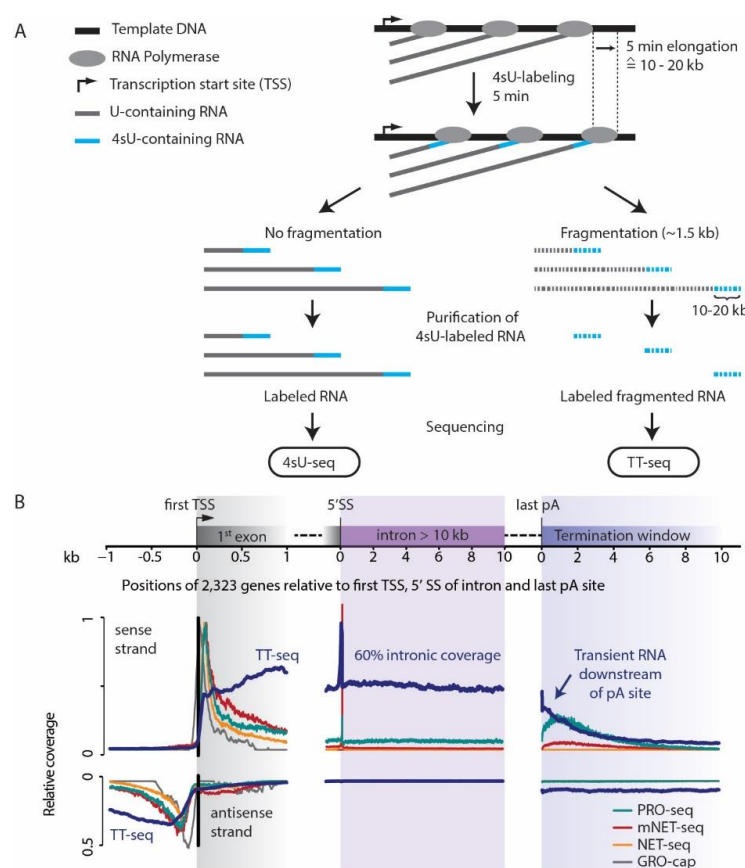


Figure 10: TT-seq enables uniform mapping of the human transient transcriptome. **(A)** Workflow of 4sU-seq and TT-seq protocols (4sU, 4-thiouridine). **(B)** Metagenome coverage comparing TT-seq to other transcriptomic methods. Average coverage in 2,323 TUs lacking paused and active genes (Core et al, 2008a) around the first TSS (left), the 5'-splice site (SS) of intronic sequences > 10 kbp (first intron excluded), and the last pA site. Signals are relative to the maximum in the first kb from the first TSS.

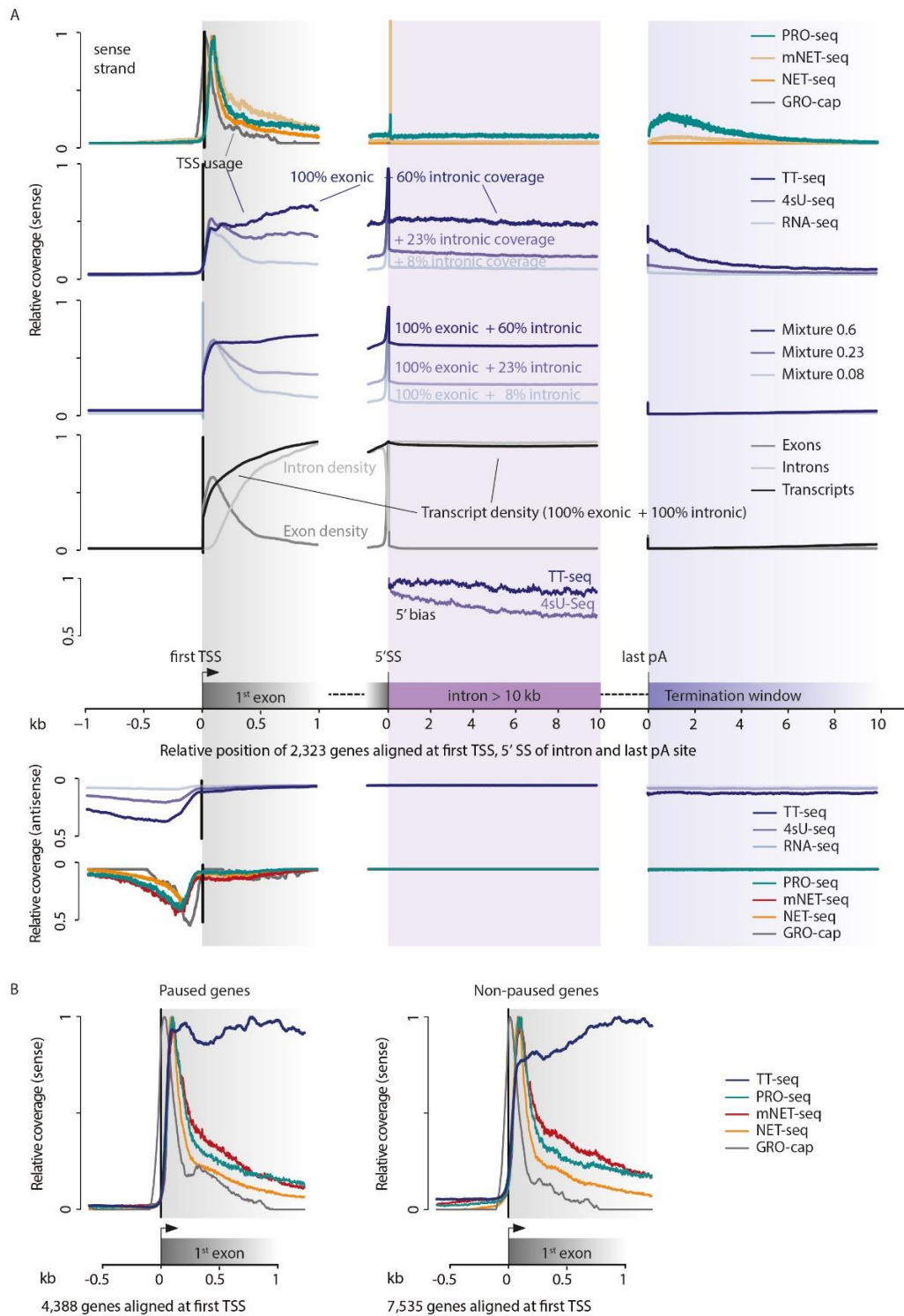


Figure 11: TT-seq maps entire transcripts. (A) Metagene analysis comparing TT-seq of 2,323 RNAs to different experimental methods and to theoretical exon, intron and mixture densities. The theoretical densities depict 8% (RNA-seq), 23% (4sU-seq) and 60% (TT-seq) intron coverage. Average coverage relative to the maximum in the first kb (left) is shown around the first TSS (left), the 5'-splice site (SS) of an intron of at least 10 kb (first intron excluded, middle), and the last pA site (right) for the sense (top panel) and antisense (bottom panel) strand or relative to the maximum in the first kb from the 5' SS (5'-bias, top panel). **(B)** Metagene analysis around the first TSS comparing TT-seq to different experimental methods for 4,388 genes classified as Class II paused genes in Core et al, 2008a (left) and for the complementary set of 7,535 non-paused genes (right).

We developed transient transcriptome sequencing (TT-seq), a protocol that maps transcriptionally active regions and enables estimation of RNA synthesis and degradation rates. TT-seq is based on 4sU-seq, which involves a brief exposure of cells to the nucleoside analog 4-thioridine (4sU; Figure 10A; Section 2.2.3; Cleary et al, 2005). 4sU is incorporated into RNA during transcription, and the resulting 4sU-labeled RNAs are isolated and sequenced. 4sU-seq is more sensitive than RNA-seq in detecting transient RNAs. However, 4sU-seq fails to map human transcripts uniformly, because only a short 3' -region of nascent transcripts is labeled during a 5-min exposure to 4sU, and the long preexisting 5' -regions dominate the sequencing data. To remove this 5' bias, TT-seq uses RNA fragmentation before isolation of labeled RNA fragments (Figure 10A; Section 2.2.3). Thus, TT-seq measures only newly transcribed RNA fragments and provides the number of polymerases transcribing a genomic position within 5 min.

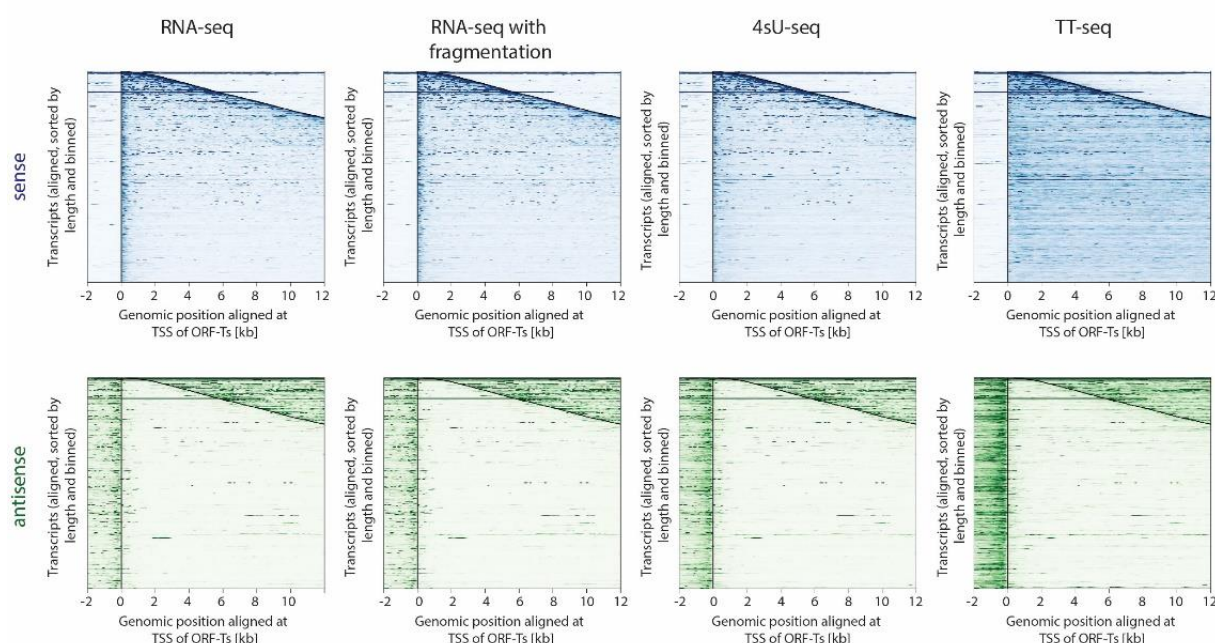


Figure 12: TT-Seq enables uniform sampling of pre-mRNAs and detects ncRNAs with high sensitivity. Coverage profiles of 2,500 most highly expressed mRNA transcripts aligned at the TSS, sorted by length and merged by averaging into horizontal bins of 10 mRNAs each for all measured samples with pooled replicates. The number of position-based read counts is color-coded, ranging from high (dark color) to low (light color). Upper and lower panels represent sense and antisense coverage, respectively.

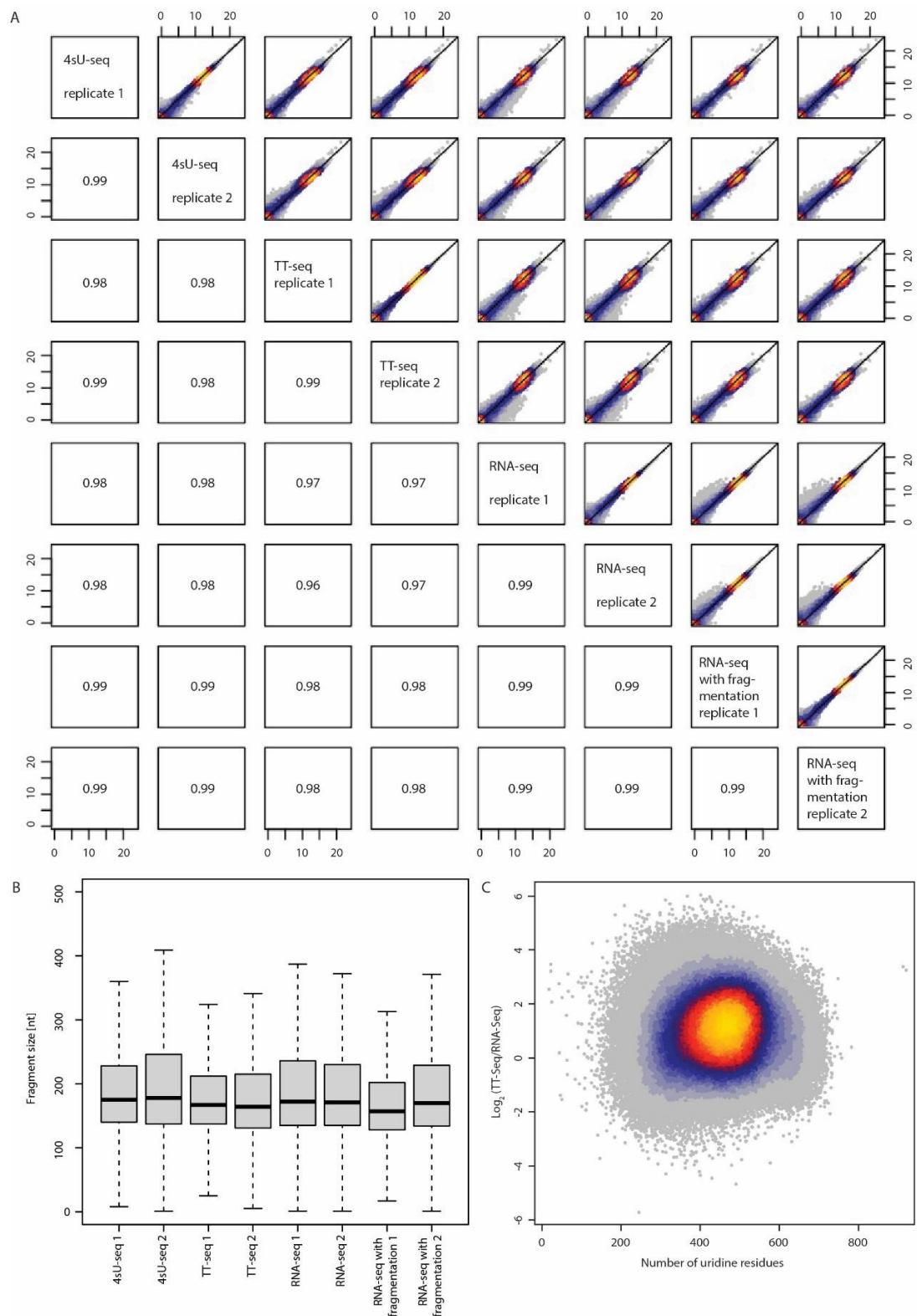


Figure 13: TT-Seq is highly reproducible, and RNA fragmentation does neither alter total RNA levels nor introduce a labeling bias. **(A)** Scatter plots comparing all measured samples in a pairwise fashion (upper triangle) and corresponding Spearman correlation (lower triangle). **(B)** Fragment size distributions for all measured samples. **(C)** Scatter plot comparing the number of uridine residues in bins of 1.5 kb against the respective log-ratio of TT-Seq versus RNA-Seq with fragmentation.

When applied to human K562 cells, TT-seq samples newly transcribed regions uniformly, whereas 4sU-seq produces a 5' bias (Figure 11A). The coverage of short-lived introns with respect to exons is estimated to be 60% for TT-seq, whereas it is 23% and 8% for 4sU-seq and RNA-seq, respectively (Figure 11A and Figure 12). TT-seq is highly reproducible (Figure 13) and enables complete mapping of transcribed regions, complementing the GRO-cap (Core et al, 2014) and CAGE (Kodzius et al, 2006) protocols, which detect RNA 5' ends (Figure 10B). TT-seq monitors RNA synthesis, whereas protocols such as PRO-seq (Kwak et al, 2013), NET-seq (Mayer et al, 2015), and mNET-seq (Nojima et al, 2015) detect RNAs attached to polymerase. Therefore, the latter protocols yield peak signals near the promoter where polymerase pauses (Figure 10B), whereas TT-seq does not. For paused and active genes (Core et al, 2008a), TT-seq reveals higher rates of RNA synthesis near the promoter (Figure 11B).

Using TT-seq data and the segmentation algorithm GenoSTAN (Zacher et al, 2016) we identified 21,874 genomic intervals of apparently uninterrupted transcription (transcriptional units, TUs; Figure 14, Figure 15A). TT-seq is highly sensitive, recovering 65% of transcription start sites (TSSs) obtained by GRO-cap (overlapping annotations within ± 400 bp; Core et al, 2014). A total of 8,543 TUs overlapped GENCODE annotations (Harrow et al, 2012), in the sense direction of transcription (50% reciprocal overlap of annotated regions; Figure 15B). This analysis detected 7,810 mRNAs, 302 long intergenic noncoding RNAs (lncRNAs), and 431 antisense RNAs (asRNAs). The 2,916 TUs that shared less than 50% of their length with GENCODE annotations were not classified. The remaining 10,415 TUs (48%) represented newly detected ncRNAs that we characterized further.

Transcripts arise from promoters but also from enhancers, which are regulatory elements with characteristic chromatin modifications (Djebali et al, 2012, Kim et al, 2010b). To detect chromatin regions comprising putative enhancers and promoters (chromatin states), we applied GenoSTAN (Zacher et al, 2016) to ENCODE ChIP-seq data (Consortium, 2012), for the coactivator p300 and a series of histone modifications (H3K27me3, H3K36me3, H4K20me1, H3K4me1, H3K4me3, H3K9ac, and H3K27ac), and to deoxyribonuclease I hypersensitivity data (Figure 16A). Of the resulting strong enhancer state regions, 81% overlapped at least one TSS from GRO-cap (Core et al, 2014), and 68% overlapped a polymerase II (Pol II) peak (Consortium, 2012), compared with 52% and 37%, respectively, for ENCODE enhancer states (Figure 16B-D).

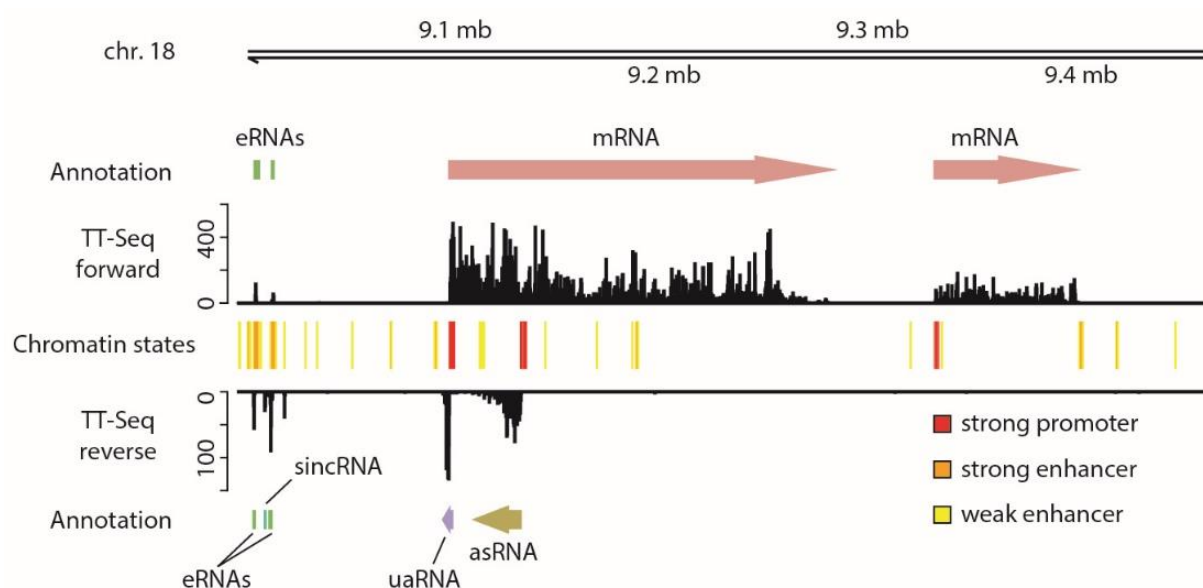


Figure 14: Example genome browser view showing RNAs from 5 of 7 transcript classes and 3 of 18 chromatin states (chr.18: 9.00 – 9.53; Mbp, million base pairs). Arrows indicate direction of transcription. Units on the y axes are read counts per 200 bp bin.

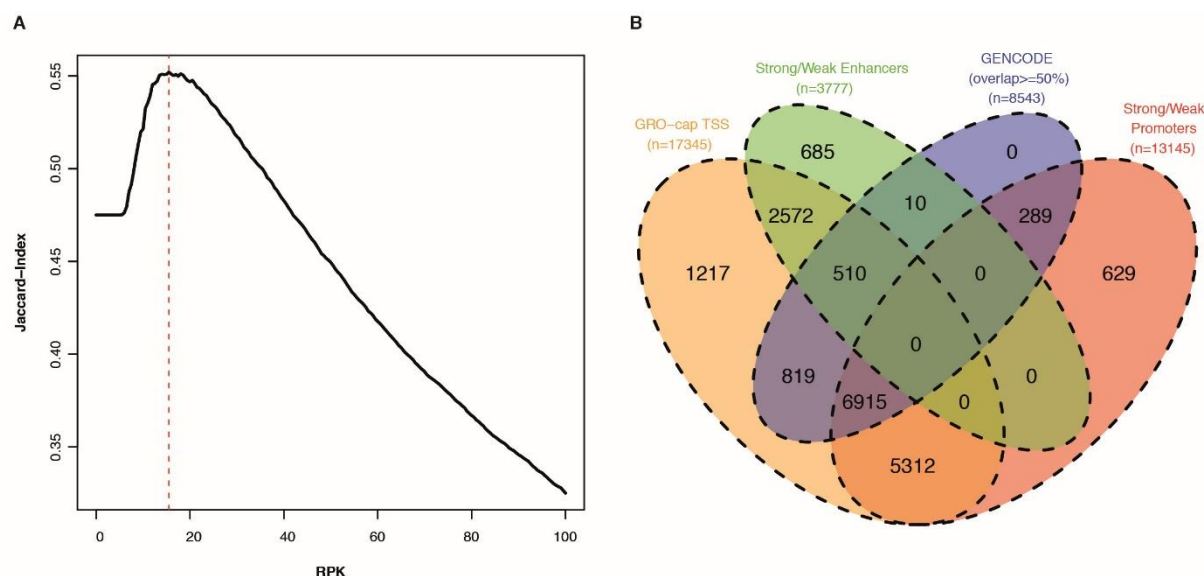


Figure 15: Accurate annotation of transcripts based on TT-Seq data using STAN (A) Jaccard index (compared to GENCODE annotation) for different choices of thresholds (RPK, x-axis). (B) Venn diagram showing the overlap of the predicted and filtered 21,874 TUs with external data sets.

The 10,415 non-annotated TUs were classified based on GenoSTAN-derived chromatin states and their positions relative to known GENCODE annotations (Figure 17A). TUs within 1 kilo-base pair (kbp) of a GENCODE mRNA TSS included 685 upstream antisense RNAs (uaRNAs; Flynn et al, 2011), and 778 convergent RNAs (conRNAs; Mayer et al, 2015). The 3,115 TUs on the strand opposite an mRNA were classified as asRNAs when they were more

than 1 kbp away from the GENCODE TSS. Remaining TUs were grouped according to their GenoSTAN chromatin state at their TSS. The 2,580 TUs that originated from promoter state regions were classified as short intergenic ncRNAs (sincRNAs; Figure 17B). Most sincRNAs (67%) were located within 10 kbp of a GENCODE mRNA TSS. The remaining 3,257 TUs originated from enhancer state regions and were classified as eRNAs (Djebali et al, 2012; Kim et al, 2010b). The newly mapped ncRNAs are short (Figure 17C). On average, lncRNAs are five times as long as sincRNAs and eRNAs have a median length of ~1000 nucleotides.

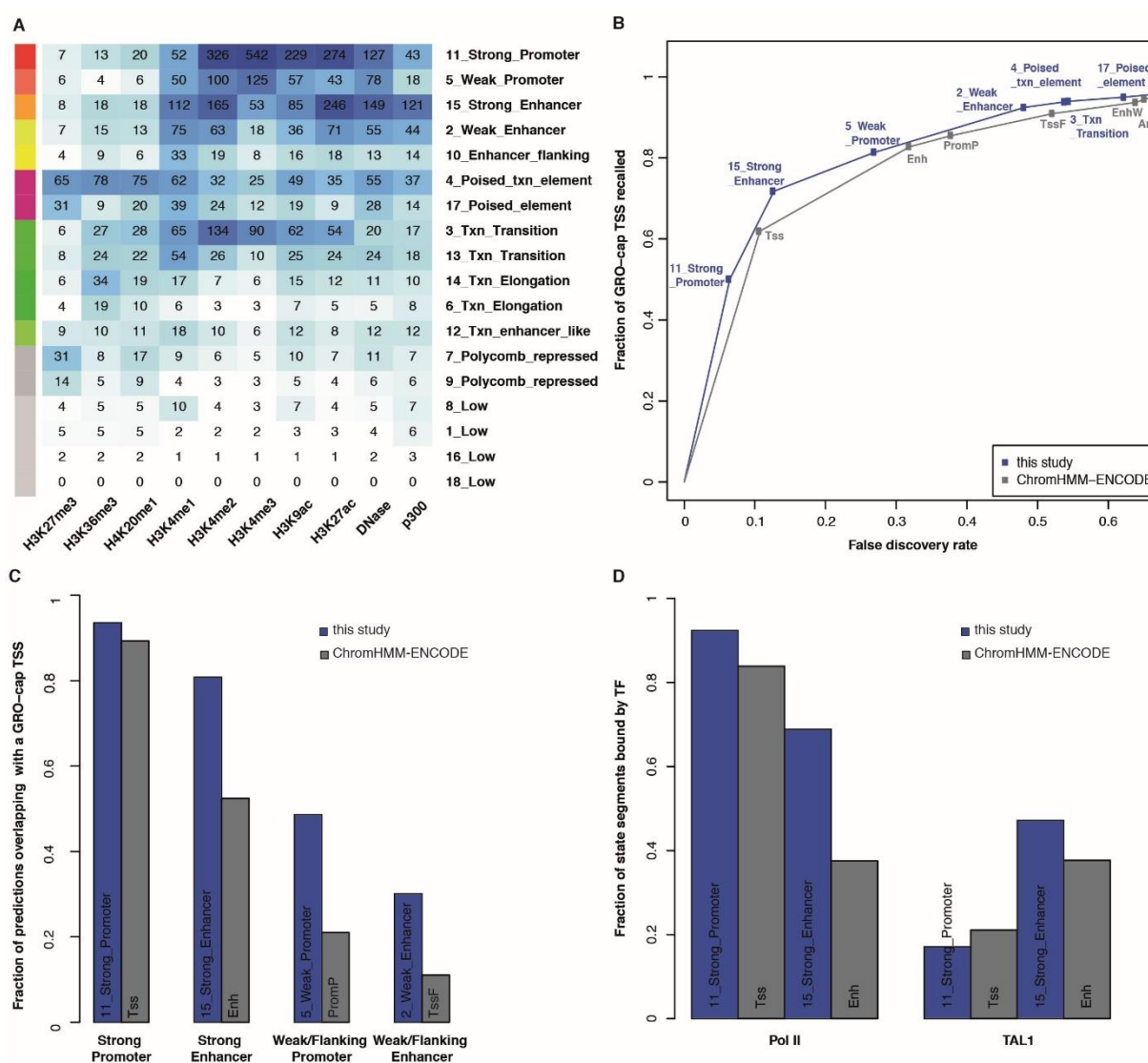


Figure 16: Chromatin states and transcriptional activity of the corresponding genomic regions. **(A)** Mean read counts for each chromatin mark and state (Txn, Transcription). **(B)** The cumulative FDR is plotted against the recall for this study and ChromHMM-ENCODE. **(C)** Fraction of recovered (overlap) GRO-cap TSSs in 4 different states (strong promoter, strong enhancer, weak/flanking promoter and weak/flanking enhancer) in this study and in ChromHMM-ENCODE. **(D)** Fraction of (promoter and enhancer) chromatin states bound by Pol II and TAL1 are shown for the annotation in this study and the ENCODE ChromHMM annotation.

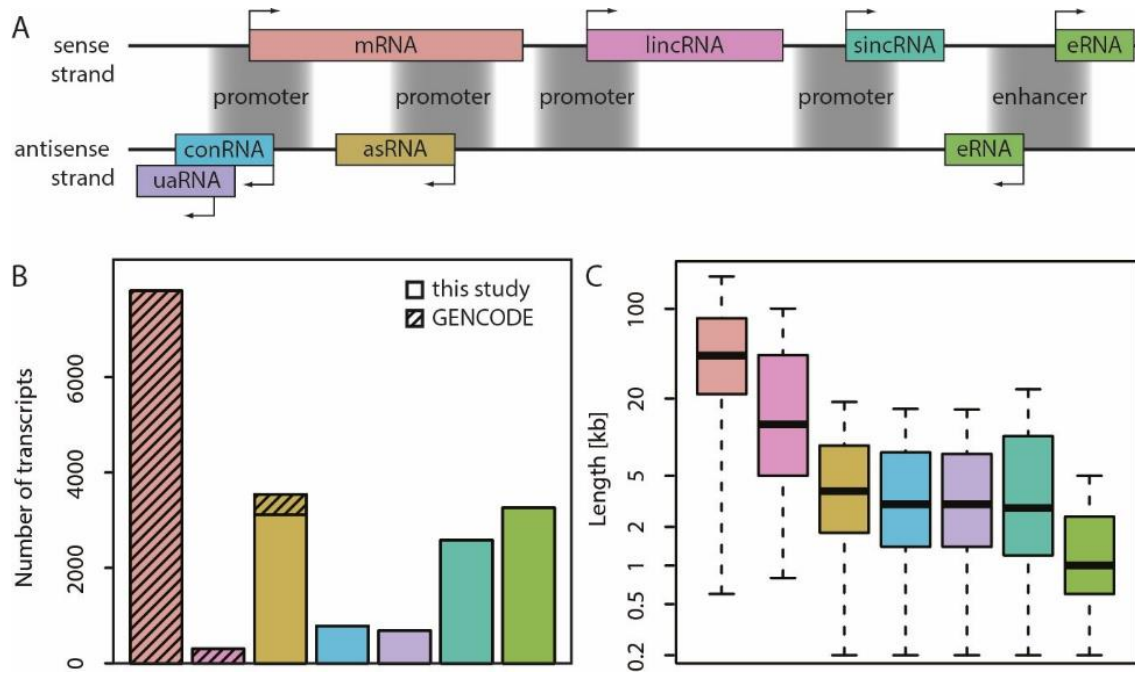


Figure 17: Annotation and lengths of RNAs mapped by TT-seq. **(A)** Definition and color code of seven transcript classes. **(B)** Number of transcripts in different classes (portions covered by GENCODE hatched). **(C)** Distribution of transcript lengths.

Kinetic modeling of TT-seq and RNA-seq data enabled us to estimate rates of RNA synthesis and degradation (Figure 18A-B). We estimated rates of phosphodiester bond formation or breakage at each transcribed position and averaged these within TUs, thus obtaining estimates of relative transcription rates and RNA stabilities. We found that RNAs and lncRNAs had the highest synthesis rates and longest half-lives. We determined a median mRNA half-life of ~50 min, compared to a previous estimate of ~139 min (Rabani et al, 2014). Other transcript classes had low synthesis rates and short half-lives, explaining why short ncRNAs are difficult to detect. eRNAs had half-lives of a few minutes, consistent with prior data (Rabani et al, 2014). Short RNA half-lives correlated with a lack of secondary structure (Figure 18C). The folding energy of eRNAs was comparable to the genomic background level (Figure 18D), and only 10% of their sequence was predicted to be structured, compared with 52% in mRNAs (Figure 18E).

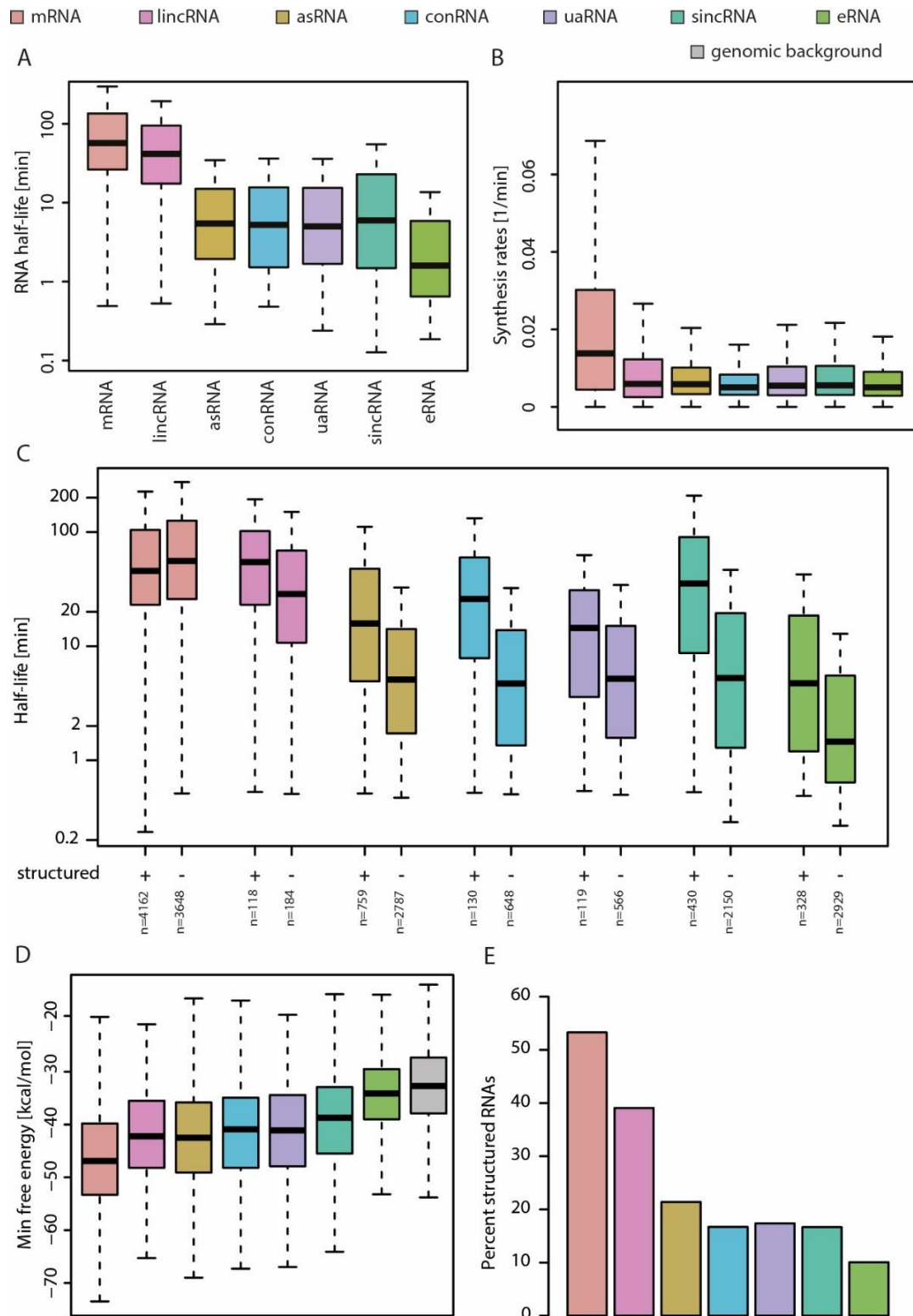


Figure 18: Estimated transcript synthesis rates, half-lives, and predicted RNA structure. **(A)** Estimated RNA half-lives for different transcript classes. Black bars represent the median, boxes represent upper and lower quartile, and whiskers represent 1.5 times the interquartile range. **(B)** Distribution of synthesis rates per transcript class. **(C)** Distribution of half-lives of different transcript classes depending on whether they are predicted to be structured or not (+, -; Washietl et al, 2005). **(D)** Distribution of the minimum free energy in the first 1000 nucleotides per transcript class. **(E)** Distribution of percentage of structured RNA in different transcript classes.

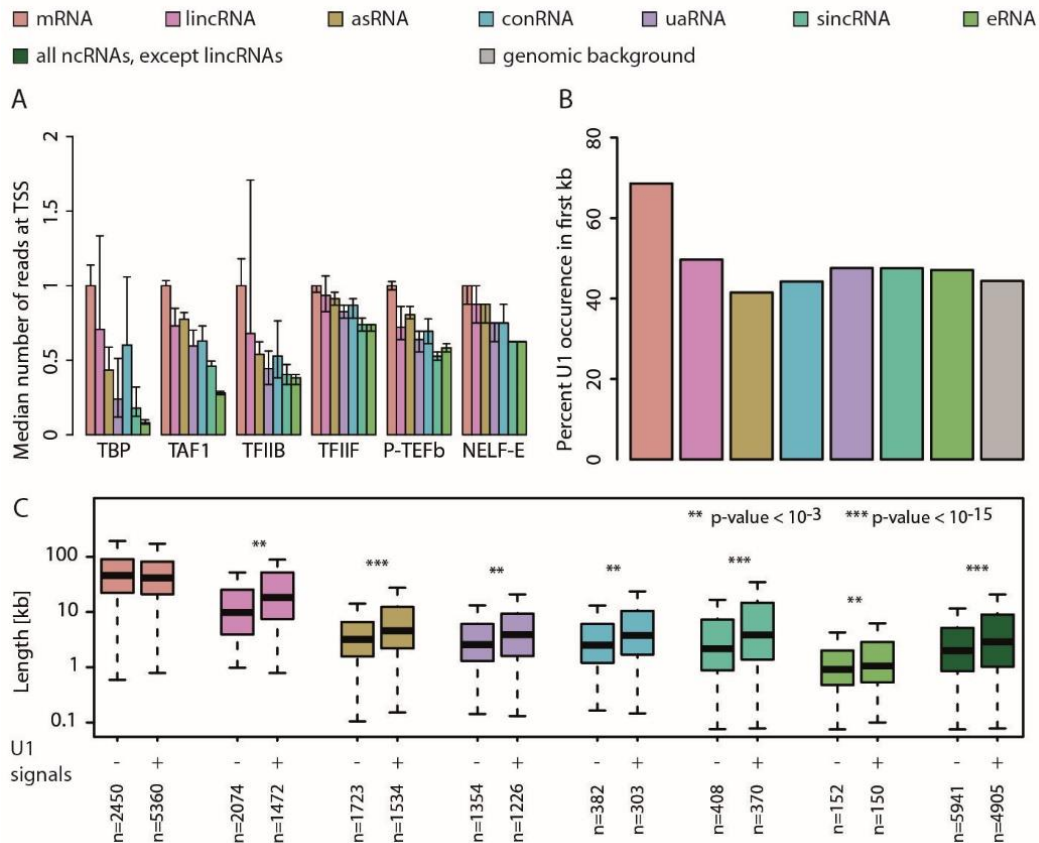


Figure 19: RNA sequence features. (A) Distribution of relative peak occupancies with factors binding promoters (ENCODE, +/- 100 bp from TSS) for transcript classes. (B) Occurrence of U1 signal in the first 1000 nt for different transcript classes. (C) Distribution of transcript lengths in transcript classes depends on the presence of U1 signals in the first 1000 nt.

We further found differences in transcription from promoters versus from enhancers (Andersson et al, 2014). Enhancers showed lower occupancy of initiation factors TBP and TAF1 than mRNA promoters did (12-fold and 3.5-fold less, respectively; $P < 10^{-16}$, Fisher's exact test), whereas TFIIB and TFIIF had similar occupancies in enhancers and promoters. Occupancies were also similar for factors involved in polymerase pausing, such as NELF-E and the P-TEFb subunit cyclin T2 (Figure 19A). Synthesis of eRNAs terminated early (Figure 17C), probably because eRNAs are not enriched in U1 small nuclear ribonucleoprotein-binding sites (U1 signals; GGUAAG, GUGAGU, or GGUGAG), that can counteract early termination and lead to RNA stabilization (Berg et al, 2012; Kaida et al, 2010). eRNAs contained U1 signals at the genomic background level (47%), whereas mRNAs were enriched (69%; $P < 10^{-16}$, Figure 19B). In all transcript classes, longer RNAs were enriched with U1 signals in the first 1000 nucleotides (Figure 19C), suggesting that evolution of stable RNAs generally involves acquisition of U1 signals.

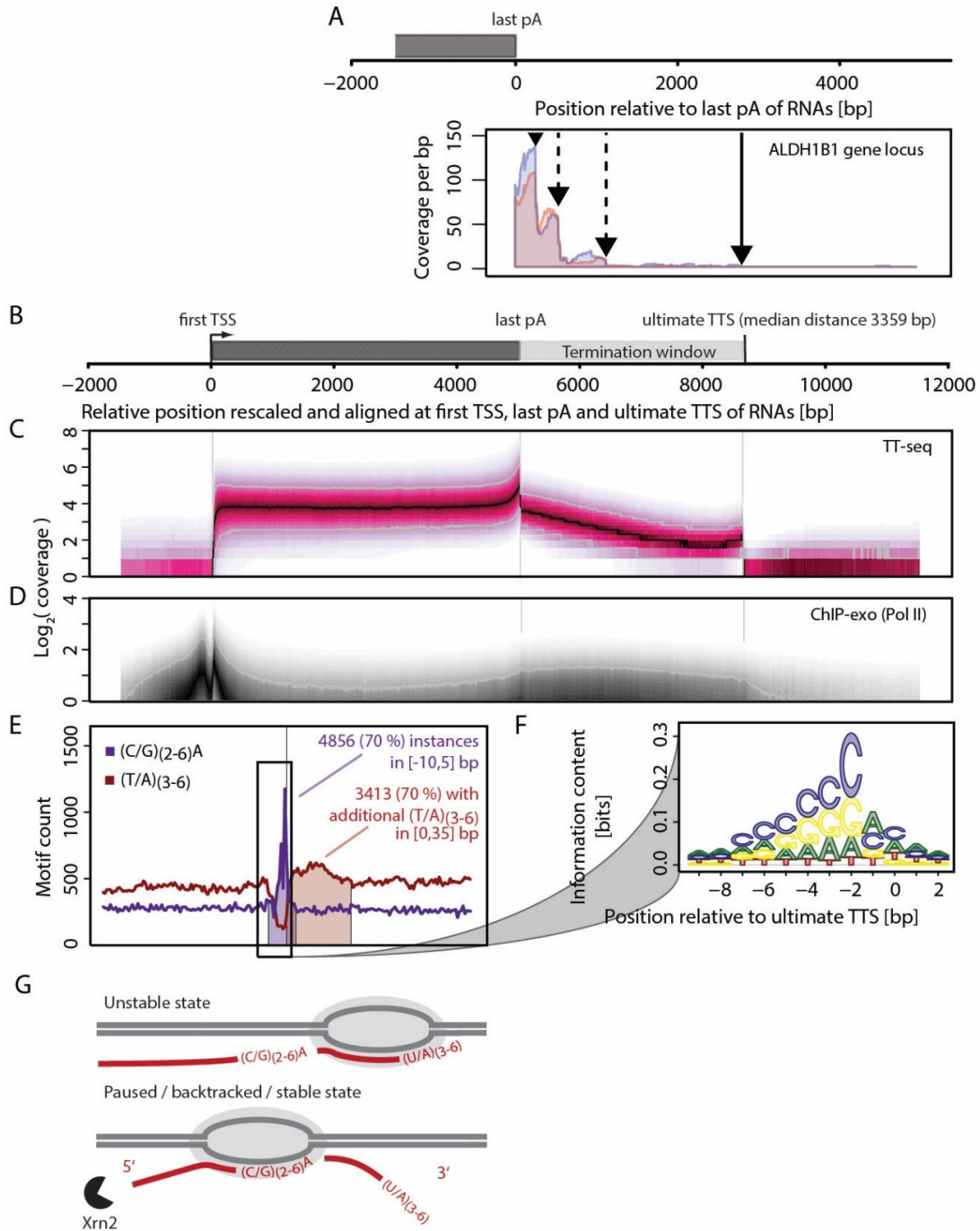


Figure 20: Transcription Termination Sites (TTSs). **(A)** TT-seq coverage for two replicates (red and blue) downstream of the pA site in the *ALDH1B1* gene locus. Arrows indicate TTSs obtained from segmentation (solid arrow, ultimate TTS). The last annotated pA site (per GENCODE) was aligned at zero. **(B)** Generic gene architecture. The first TSS was aligned at zero, and the last pA site was set at a rescaled distance of 5,000 bp from the TSS (the real median distance is 24,079 bp for 6,977 investigated genes). The ultimate TTS is depicted at a median distance of 3,359 bp from the last pA site (rescaled). **(C)** TT-seq coverage over quantiles 0.05 to 0.95 (pink area; black center line, median), rescaled and aligned as in (B). **(D)** Pol II occupancy determined by the ChIP-exonuclease method (Venters & Pugh, 2013), visualized as in (C) (white center line, upper quartile). **(E)** (C/G)₍₂₋₆₎A and (T/A)₍₃₋₆₎ sequence count within ± 100 bp of ultimate TTSs. **(F)** PWM (position weight matrix) logo representation of nucleotides at positions -9 to $+2$ around the ultimate TTS (position 0). **(G)** Predicted polymerase states at the T-rich stretch downstream of the TTS (top) and after backtracking to the TTS (bottom) (gray, DNA; red, RNA).

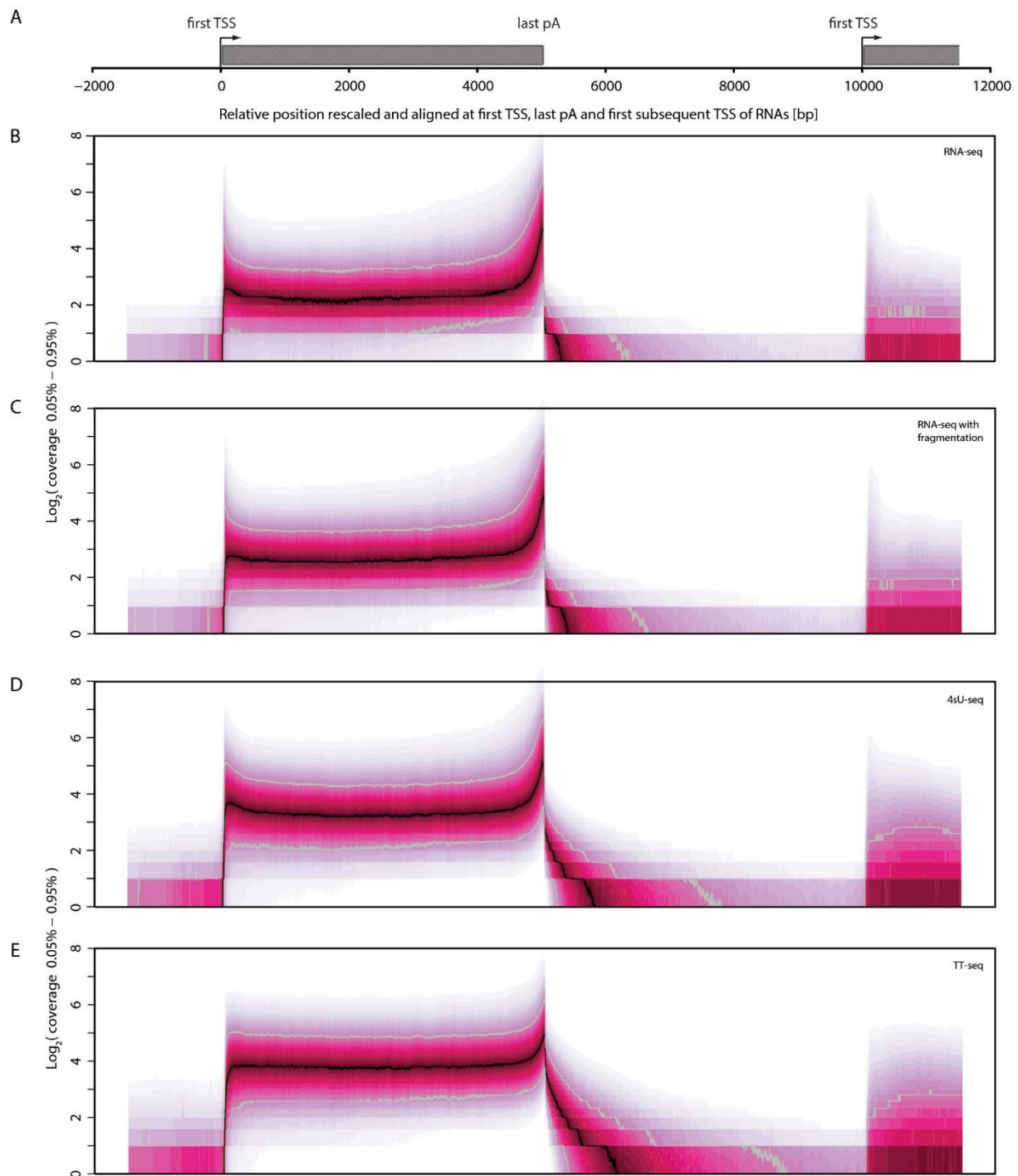


Figure 21: Comparison of RNA synthesis downstream of the pA site in different RNA-Seq experiments. **(A)** Generic gene architecture. Genomic position of the first TSS is aligned at 0. Last pA site is located at a distance of 5,000 bp from the first TSS for visualization purposes instead of a median of 24,079 bp for 6,977 investigated genes. The subsequent first TSS is depicted at a median distance of additional 5,000 bp from the last pA site. **(B)** RNA-Seq coverage fanned out over 0.05% - 0.95% quantile range, rescaled and aligned according to schematic in (A) in logarithmic scale for 6,977 investigated genes. **(C)** RNA-Seq with fragmentation coverage as in (B). **(D)** 4sU-Seq coverage as in (B). **(E)** TT-Seq coverage as in (B).

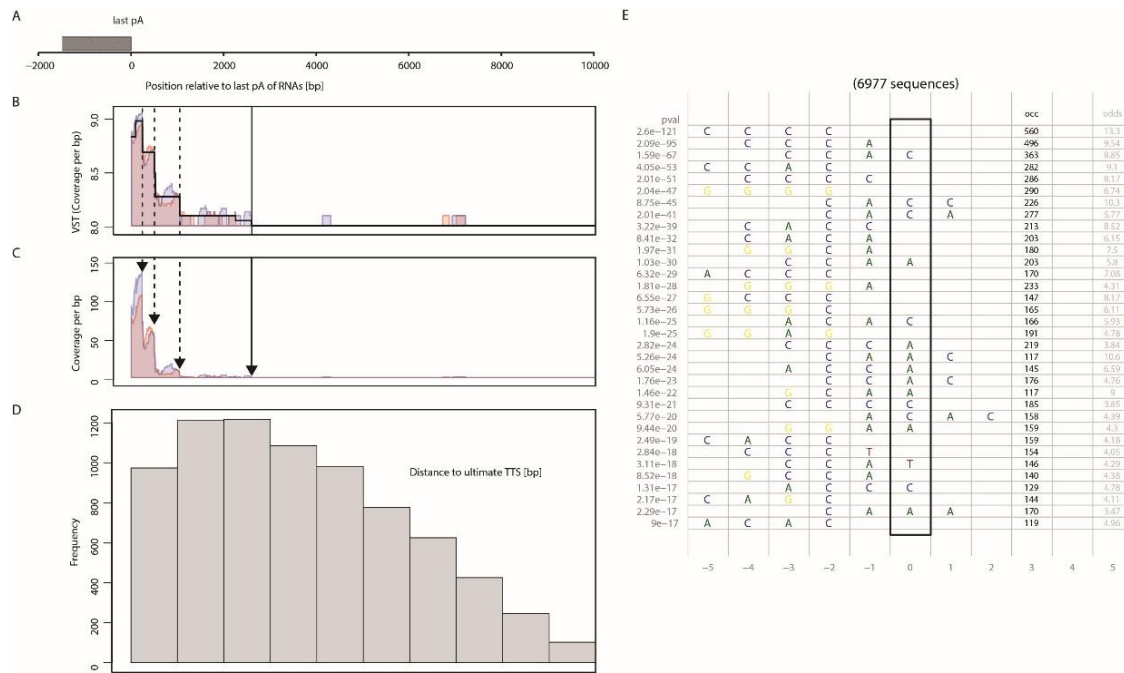


Figure 22: Estimation of ultimate TTSs via segmentation of potential termination window. (A) Schematic of generic gene architecture showing the last annotated pA site (GENCODE). Genomic position of the last pA site is aligned at 0. The window for site determination is extended to the subsequent first TSS or max. 10 kb. (B) Variance stabilization transformed (VST) TT-seq coverage for two replicates (red, blue) at the ALDH1B1 gene loci. The black line depicts the optimal stepwise linear function estimated via a segmentation algorithm; the breakpoints are indicated by the dashed lines. (C) Plot shows the same loci as in (B) for non-VST normalized data. Dashed and solid arrows indicate putative and ultimate TTSs resulting from the segmentation depicted in (B). (D) Distribution of estimated TTSs relative to last pA site for 6,977 investigated genes. (E) Plot showing the top 35 enriched 4-mers found by comparing the frequency of all possible 4-mers in a window of +/- 5 bp around the estimated TTS for fixed positions. Testing was done via Fisher's exact tests against the (background) frequency of the respective 4-mer obtained from a window of the same size shifted 30 bp downstream. The respective p-values and odd-ratios are given in the left and right panel.

TT-seq also enabled us to uncover transcription termination sites (TTSs). TT-seq detected transient RNA downstream of the polyadenylation (pA) site (Figure 10B, Figure 20, Figure 21). Such RNA is difficult to detect, because RNA cleavage at the pA site leads to an unprotected 5' end and RNA degradation by the XRN2 exonuclease (Almada et al, 2013; Kim et al, 2004; West et al, 2004). For a total of 6,977 mRNA genes, we derived, on average, four TTSs (Figure 22, Figure 23). TTSs were located within a termination window that extended from the last pA site to an "ultimate TTS," where RNA coverage dropped to background levels (Figure 20A-C). The termination window had a median width of ~3,300 bp and could be up to 10 kbp wide (Figure 22D), consistent with Pol II occupancy data (Figure 20D; Venters & Pugh, 2013). For the 5,113 TTSs with the strongest drop in TT-seq signal, Pol II peaks were obtained by PRO-seq (Kwak et al, 2013), NET-seq (Mayer et al, 2015), and mNET-seq (Figure 24; Nojima et al, 2015), indicating that Pol II pauses at the TTS.

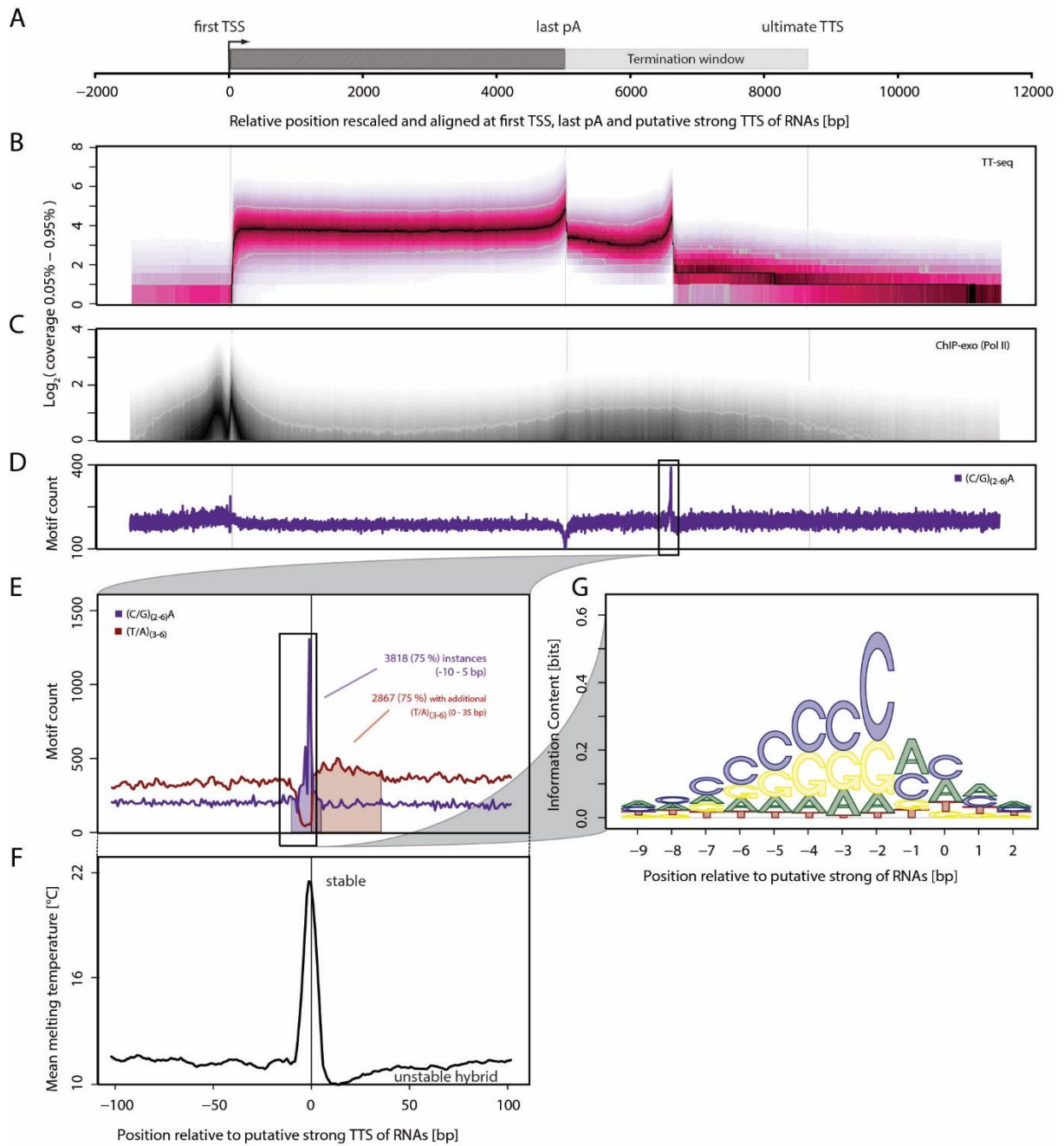


Figure 23: Global detection and nature of TTSs inside the termination window. **(A)** Generic gene architecture. Genomic position of the first TSS is aligned at 0. The last pA site is located at a distance of 5,000 bp from TSS for visualization purposes for 5,113 investigated genes. The estimated ultimate TTS is depicted at a median distance of 3,359 bp from the last pA site. **(B)** TT-Seq coverage fanned out over 0.05% - 0.95% quantiles, rescaled and aligned according to (A) in log-scale for 5,113 investigated genes with the exception that the putative strong TTS is used in a median distance of 1,593 bp from the last pA. **(C)** ChIP-exo Pol II occupancy (Venters & Pugh, 2013) as in (B). **(D)** (C/G)₍₂₋₆₎A kmer count in the corresponding RNA sequence rescaled and aligned according to schematic in (A). **(E)** (C/G)₍₂₋₆₎A and (T/A)₍₃₋₆₎ kmer count in the corresponding RNA sequence in a window of +/- 100 bp around estimated putative strong TTS. **(F)** The mean melting temperature for a window of +/-100 bp around the estimated putative strong TTS. **(G)** PWM logo representation of -9 to +2 bp of the corresponding RNA sequence around the putative strong TTS (position 0).

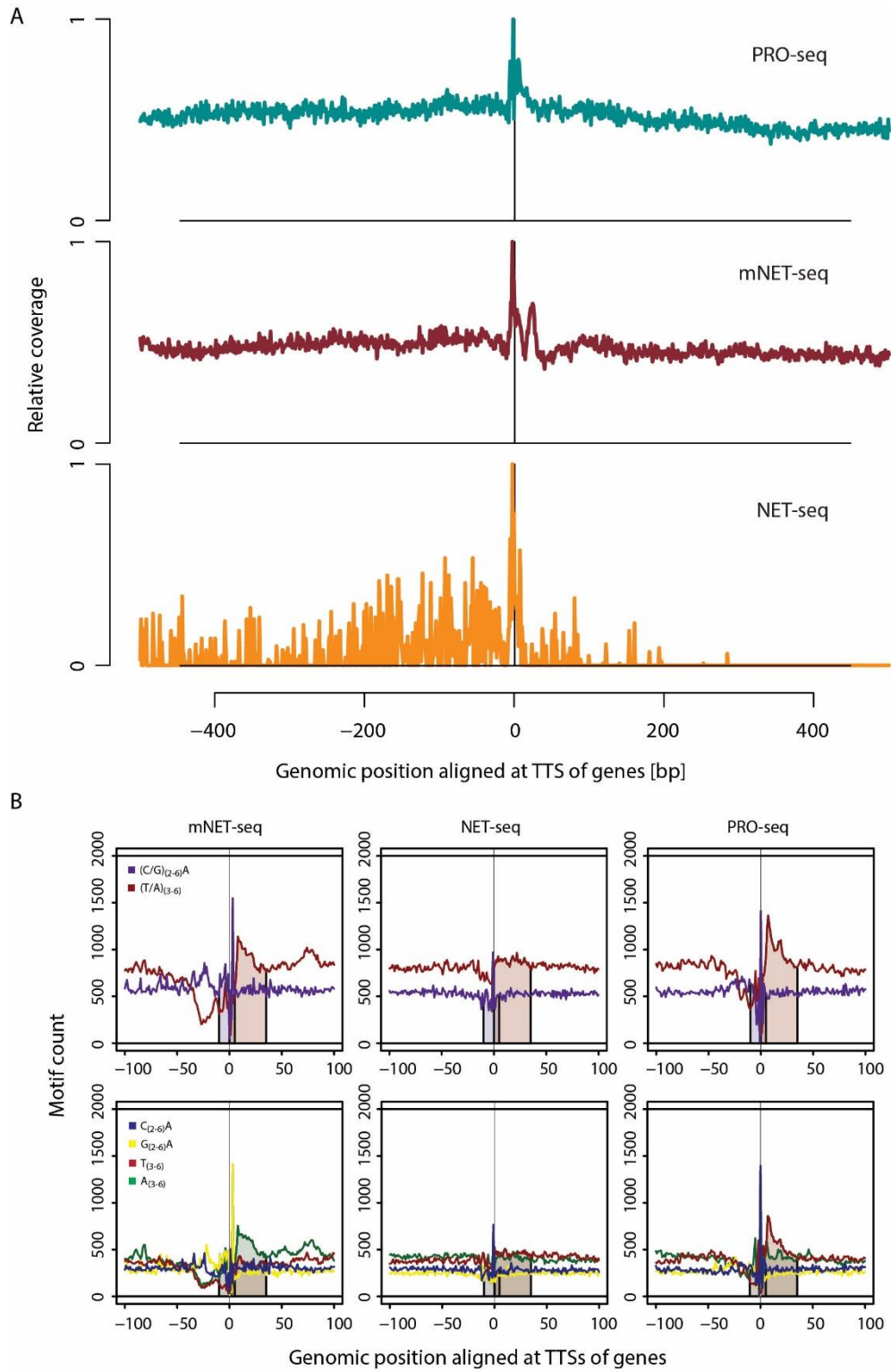


Figure 24: The derived TTSs coincide with sites of polymerase pausing. **(A)** Average PRO-seq, mNET-seq and NET-seq signal is shown around the putative strong TTSs for 5,113 investigated genes relative to maximum. **(B)** (C/G)₍₂₋₆₎A and (T/A)₍₃₋₆₎ kmer counts in the corresponding RNA sequence underlying putative strong TTSs derived from PRO-seq, mNET-seq and NET-seq for 14,060 investigated genes in a window of +/- 100 bp (the TTS was set to be the position with the maximum number of 3' end read counts for all three methods between the last pA site and the next annotated downstream feature). Lower panel shows C₍₂₋₆₎A, G₍₂₋₆₎A, T₍₃₋₆₎ and A₍₃₋₆₎ kmer counts depicted analogously.

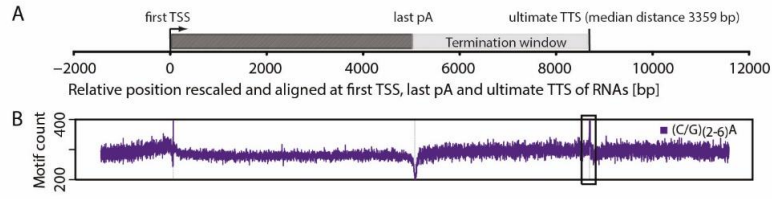


Figure 25: Distribution of (C/G)₍₂₋₆₎A kmers. **(A)** Generic gene architecture. The first TSS was aligned at zero, the last pA site was set at a rescaled distance of 5,000 bp from TSS (real median distance is 24,079 bp for 6,977 investigated genes). The ultimate TTS is depicted at a median distance of 3,359 bp from the last pA site (rescaled). **(B)** (C/G)₍₂₋₆₎A kmer sequence count rescaled and aligned as in (A).

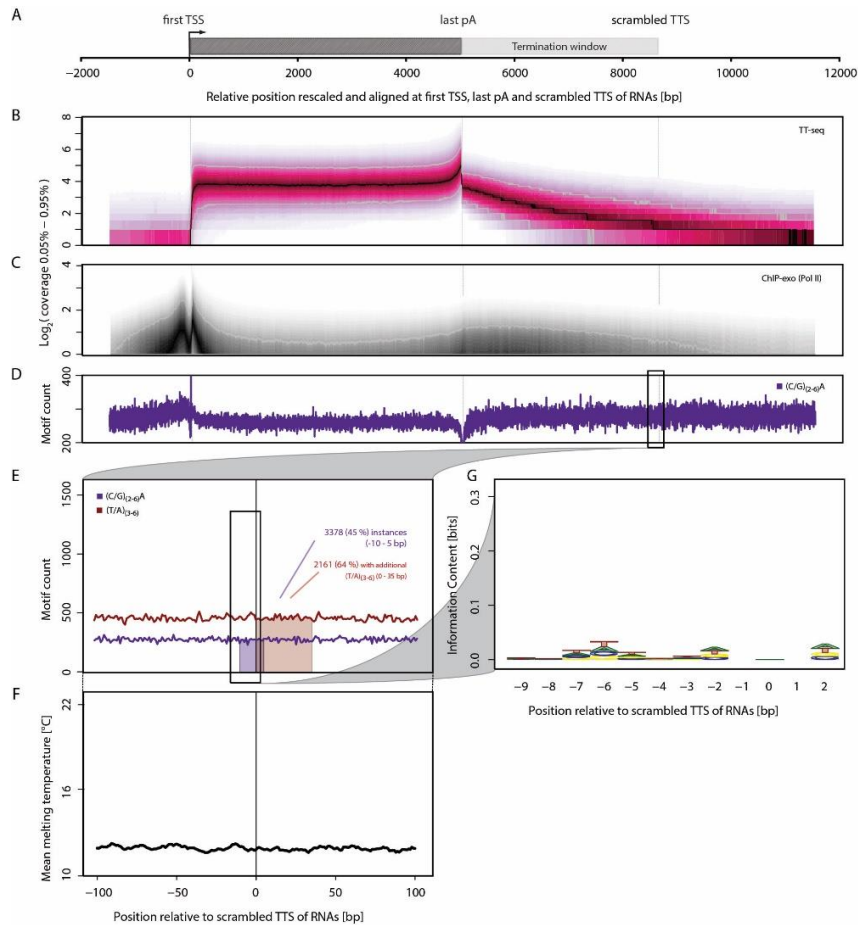


Figure 26: Supporting evidence for derived TTSs. **(A)** Schematic of generic gene architecture. Genomic position of the first TSS is aligned at 0. The last pA site is located at a distance of 5,000 bp from the first TSS for visualization purposes instead of a median of 24,079 bp for 6,977 investigated genes. The estimated ultimate TTS is depicted at a median distance of 3,359 bp from the last pA site. **(B)** TT-Seq coverage fanned out over 0.05% - 0.95% quantile range rescaled and aligned according to schematic in (A) in logarithmic scale for 6,977 investigated genes with the exception that the estimated distances from the last pA to the ultimate TTS are shuffled across all 6,977 genes. **(C)** ChIP-exo (Pol II) coverage (Venters & Pugh, 2013) as in (B). **(D)** (C/G)₍₂₋₆₎A kmer count in the corresponding RNA sequence rescaled and aligned according to schematic in (A). **(E)** (C/G)₍₂₋₆₎A and (T/A)₍₃₋₆₎ kmer count in the corresponding RNA sequence in a window of +/- 100 bp around estimated shuffled ultimate TTS. **(F)** The mean melting temperature for a window of +/-100 bp around the estimated shuffled ultimate TTS. **(G)** PWM logo representation of -9 to +2 bp of the corresponding RNA sequence around the shuffled ultimate TTS (position 0).

The derived TTSs are strongly enriched for the sequence (C/G)₍₂₋₆₎A (window ± 5 bp; $P < 10^{-16}$, Fisher's exact test; odds ratios, 2.98 and 1.63 for C₃A and G₃A, respectively; Figure 20E and Figure 25, Figure 26). This sequence can contain up to six cytosines or guanines (Figure 20F and Figure 22E). A G₃A element has also been found in a known termination signal (Ashfield et al, 1994). The C₍₂₋₆₎A and G₍₂₋₆₎A sequences are generally followed by a T-rich [T₍₃₋₆₎] or an A-rich stretch [A₍₃₋₆₎], respectively ($P < 10^{-16}$; odds ratios, 2.39 and 1.24 for T₄ and A₄, respectively), that are located, on average, 15 bp downstream of the TTS (Figure 20E). Such sequences were found at TTSs of all TUs (Figure 27) and can even be derived from published data (Figure 24B). In summary, the detected TTSs were highly enriched with the consensus motif (C/G)₍₂₋₆₎AN_x(T/A)₍₃₋₆₎, where N_x is a short stretch of nucleotides.

To test for the in vivo functionality of the derived TTS motif, we transfected expression plasmids into K562 cells that either lacked or contained four C₃AN₈T₄ or C₇AN₈T₄ motifs within 600 bp downstream of the pA site (Figure 28A, Table 9, Table 10, Table 15, Table 16, Section 2.2.5). When the TTS motifs were present, significantly less RNA was detected downstream of the motifs, indicating termination of a fraction of polymerases (Wilcoxon test; Figure 28B). This experiment supports the functionality of the derived TTS motif in vivo. Termination depended on an upstream pA signal (Figure 28B), consistent with an occurrence of the motif in gene bodies, where they do not lead to transcription termination.

Table 15: Ct-values from qPCR experiments

Samples	Transfection efficiency primers – Ct values, technical replicates				Termination read-through primers - Ct values, technical replicates			
	1	2	3	4	1	2	3	4
pA Control 1A	12.86	12.66	13.11		17.07	17.1	17.04	
pA Control 1B	12.8	12.56	12.69		17.27	17.32	17.03	
pA Control 1C	14.08	14.12	14.1		20.66	20.7	20.81	
pA C ₃ AN ₈ T ₄ 1A	13.07	13.21	13.34		18.05	18.21	18.17	
pA C ₃ AN ₈ T ₄ 1B	13.15	13.02	13.09		17.79	17.67	17.93	
pA C ₃ AN ₈ T ₄ 1C	14.33	14.28	14.35		21.37	21.37	21.44	
pA C ₇ AN ₈ T ₄ 1A	13.05	13.04	13.27		17.55	17.85	17.89	
pA C ₇ AN ₈ T ₄ 1B	13.33	13.42	13.36		18.39	18.28	18.33	
pA C ₇ AN ₈ T ₄ 1C	12.36	12.52	12.47		20.2	20.32	20.39	
No pA Control 1A	11.73	12.38	12.31	12.45	13.04	13.4	13.27	13.31
No pA Control 1B	11.64	12.53	12.32	12.26	12.92	13.23	13.34	13.77
No pA Control 2A	13.4	13.79			14.67	14.93		

No pA C ₃ AN ₈ T ₄ 1A	11.69	12.41	12.33	12.27	13.03	13.23	13.8	13.36
No pA C ₃ AN ₈ T ₄ 1B	12.01	12.64	12.6	12.69	13.02	13.27	13.3	13.46
No pA C ₃ AN ₈ T ₄ 2A	13.11	13.11			14.24	14.02		
No pA C ₇ AN ₈ T ₄ 1A	11.94	12.63	12.62	12.76	13.37	13.89	13.8	13.84
No pA C ₇ AN ₈ T ₄ 1B	11.62	12.51	12.46	12.51	12.93	13.41	13.28	13.7
No pA C ₇ AN ₈ T ₄ 2A	13.78	13.59			14.68	14.48		

Table 16: normalized values from qPCR experiment

Samples	normalized qPCR values - technical replicates			
	1	2	3	4
pA Control 1A	1	0.979	1.021	NA
pA Control 1B	0.957	0.924	1.13	NA
pA Control 1C	1.045	1.016	0.942	NA
pA C ₃ AN ₈ T ₄ 1A	0.637	0.57	0.586	NA
pA C ₃ AN ₈ T ₄ 1B	0.883	0.959	0.801	NA
pA C ₃ AN ₈ T ₄ 1C	0.744	0.744	0.709	NA
pA C ₇ AN ₈ T ₄ 1A	0.849	0.689	0.671	NA
pA C ₇ AN ₈ T ₄ 1B	0.709	0.765	0.739	NA
pA C ₇ AN ₈ T ₄ 1C	0.458	0.421	0.401	NA
No pA Control 1A	1.161	0.904	0.99	0.963
No pA Control 1B	1.315	1.061	0.983	0.73
No pA Control 2A	1.094	0.914	NA	NA
No pA C ₃ AN ₈ T ₄ 1A	1.135	0.988	0.665	0.903
No pA C ₃ AN ₈ T ₄ 1B	1.508	1.268	1.242	1.111
No pA C ₃ AN ₈ T ₄ 2A	1.053	1.227	NA	NA
No pA C ₇ AN ₈ T ₄ 1A	1.113	0.776	0.826	0.804
No pA C ₇ AN ₈ T ₄ 1B	1.388	0.995	1.089	0.814
No pA C ₇ AN ₈ T ₄ 2A	1.157	1.329	NA	NA

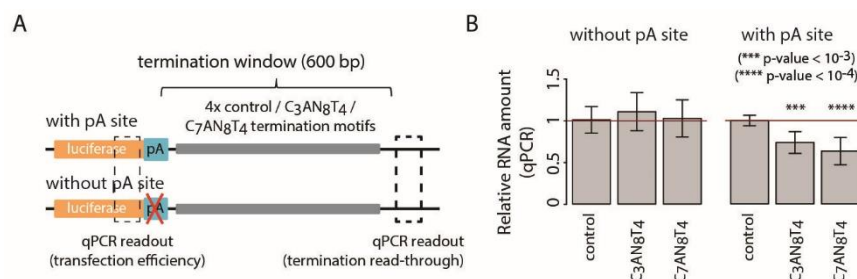


Figure 28: Experimental support for the functionality of the derived TTS motifs. **(A)** Schematic of in vivo transcription assay to test the TTS motif. **(B)** Barplot showing the relative RNA abundance (qPCR) downstream of four TTS motifs relative to a control sequence without and with pA site.

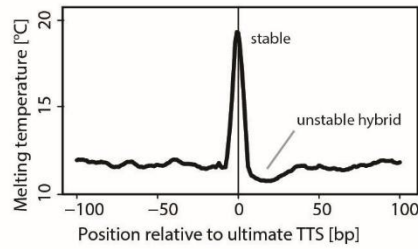


Figure 29: Distribution of hybrid stability at TTS. The mean melting temperature for each window of +/-100 bp around the estimated TTSs was calculated as the gene-wise position based estimate of the melting temperature of 8-base pair DNA-RNA hybrids.

Transcription over the $(C/G)_{(2-6)}AN_x(T/A)_{(3-6)}$ sequence is predicted to destabilize the polymerase complex (Ashfield et al, 1994; Kireeva et al, 2000), because the melting temperature of the DNA-RNA hybrid is low in the T/A-rich region (Figure 29). This may trigger backtracking and trap polymerase at the TTS (Figure 20G). At the TTS, the hybrid is C/G-rich and stable, and RNA may be cleaved from its 3' end to yield a terminal A residue. Polymerase can then be released from DNA by XRN2 (Kim et al, 2004; West et al, 2004). TT-seq has afforded insights into the determinants of human genome transcription and provides a complementary tool for transcriptome analysis.

3.2 Enhancer RNAs and their paired promoter RNAs arise simultaneously after stimulation

All results presented in this Section were obtained in collaboration with Carina Demel and are in the process of being submitted (Michel et al, 2016, *unpublished*). For detailed author contributions see page 7.

Mammalian genomes are more transcribed than previously assumed (Cheng et al, 2005; Consortium, 2012). Recent technology and method developments have led to an updated view on transcription; namely that it produces not only protein-coding transcripts (mRNAs) but also non-coding transcripts (ncRNAs) that act on various levels in cell function and identity (Dhanasekaran et al, 2013). Enhancer RNAs (eRNAs) are one class of ncRNAs (Djebali et al, 2012; Kim et al, 2010b). They emerge from enhancer regions; genomic regulatory units consisting of multiple binding sites for transcription factors that can increase mRNA transcription in an orientation-independent manner and can act from a long-range distance (Banerji et al, 1981). When activated, enhancers adopt an open chromatin structure (Calo & Wysocka, 2013), are bound by specific transcription factors and co-activators that recruit Mediator (Fan et al, 2006). Subsequently, mediator can bridge the enhancer to the promoter by binding to general transcription factors and the RNA Polymerase II (Pol II) at the promoter (Plaschka et al, 2015). Gene-enhancer looping, which is facilitated by CTCF and cohesin, increases stability of the initiation complex and promotes successful promoter escape (Allen & Taatjes, 2015; DeMare et al, 2013; Splinter et al, 2006).

Enhancers are a hallmark of gene regulation and cellular identity but identifying them and their associated promoters has been a tedious task (Shlyueva et al, 2014). Enhancers can be distinguished from other genomic regions through a distinct histone tail signature (Heintzman et al, 2007; Schubeler, 2007; Visel et al, 2009), but this method requires a large amount of available chromatin immunoprecipitation (ChIP) data sets. An easier and more direct method is to detect eRNAs since they are a good proxy for enhancer activity (Li et al, 2016; Melgar et al, 2011; Wu et al, 2014). eRNAs are not conserved (Andersson et al, 2014), transient (Schwalb et al, 2016), and rapidly targeted for degradation by the exosome (Lubas et al, 2015).

The role of enhancer transcription has been widely discussed. Recent findings suggest that transcription of the enhancer has to precede the promoter transcription, while others argue that

eRNAs are merely a by-product of promoter transcription, where the polymerase transcribes the enhancer due to the spatial proximity of promoter and enhancer (Li et al, 2016). Several studies detected a delay between the transcription of enhancers and promoters (Arner et al, 2015; De Santa et al, 2010b; Kaikkonen et al, 2013; Schaukowitch et al, 2014). Some studies show impairment of target mRNA activation after eRNA knockdown in single-gene assays (Li et al, 2013; Schaukowitch et al, 2014). Negative effects on enhancer-promoter looping after eRNA knockdown was also shown (Li et al, 2013), but not in every study (Hah et al, 2013; Schaukowitch et al, 2014). In conclusion, it remains unclear which role eRNAs carry in mRNA activation.

Here we use TT-seq to detect eRNAs and learn about enhancer and promoter activity. TT-seq is a sensitive method to detect transient RNAs and calculate synthesis and degradation rates, which is based on metabolic labeling of nascent RNAs with 4-thiouridine (4sU; Schwalb et al, 2016). Fragmentation of the labeled RNA prior to isolation increases resolution and gives a more accurate view on transcriptional activity within the labeling time. We use T cell activation as a highly dynamic and rapid cellular process and analyze the first 15 minutes after activation in 5 min intervals. T-cell activation is one of the most studied cellular responses to stimulation. Activation comprises rapid signaling cascades via protein-protein interactions, phosphorylation/dephosphorylation of target signaling molecules, and intracellular calcium release leading to a change in gene expression (Crabtree, 1989; Ellisen et al, 2001; Feske et al, 2001; Marrack et al, 2000; Raghavan et al, 2002; Rogge et al, 2000).

Taken together, we use TT-seq in this study to apply a highly sensitive method to the very rapid and dynamic system of T cell activation in order to learn about the temporal resolution of enhancer and promoter activation.

TT-seq identifies thousands of differentially expressed genes very early during T cell activation

We performed TT-seq in Jurkat cells after activation with Phorbol 12-myristate 13-acetate (PMA) and ionomycin in 5 min intervals for 15 min (Figure 30A). The time points were 0 min (before activation), 5 min, 10 min and 15 min (after activation). Already after 5 minutes, we observed strong upregulation and downregulation of protein-coding transcripts (mRNAs) in the TT-seq data compared to the Total RNA-seq data (Figure 30B and C). We also

observed an almost uniform coverage of transcripts over introns and transcription after the polyA site (PAS) annotated by GENCODE. This shows that our data is sensitive for transient RNAs (Figure 30B and C).

Next, we segmented the genome into expressed and unexpressed regions using GenoSTAN (Zacher et al, 2016), to annotate transcripts present in Jurkat cells before and after activation (Figure 31A). We annotated a total of 24,835 transcripts (Figure 31B, RPK cutoff = 16.5). By overlapping our transcripts with GENCODE (Harrow et al, 2012), we classified a total of 9,576 mRNAs, 590 lncRNAs and 14,669 non-coding RNAs (ncRNAs; Figure 31C). The length of the RNA classes correlated with previous reports (Figure 31D; Schwalb et al, 2016).

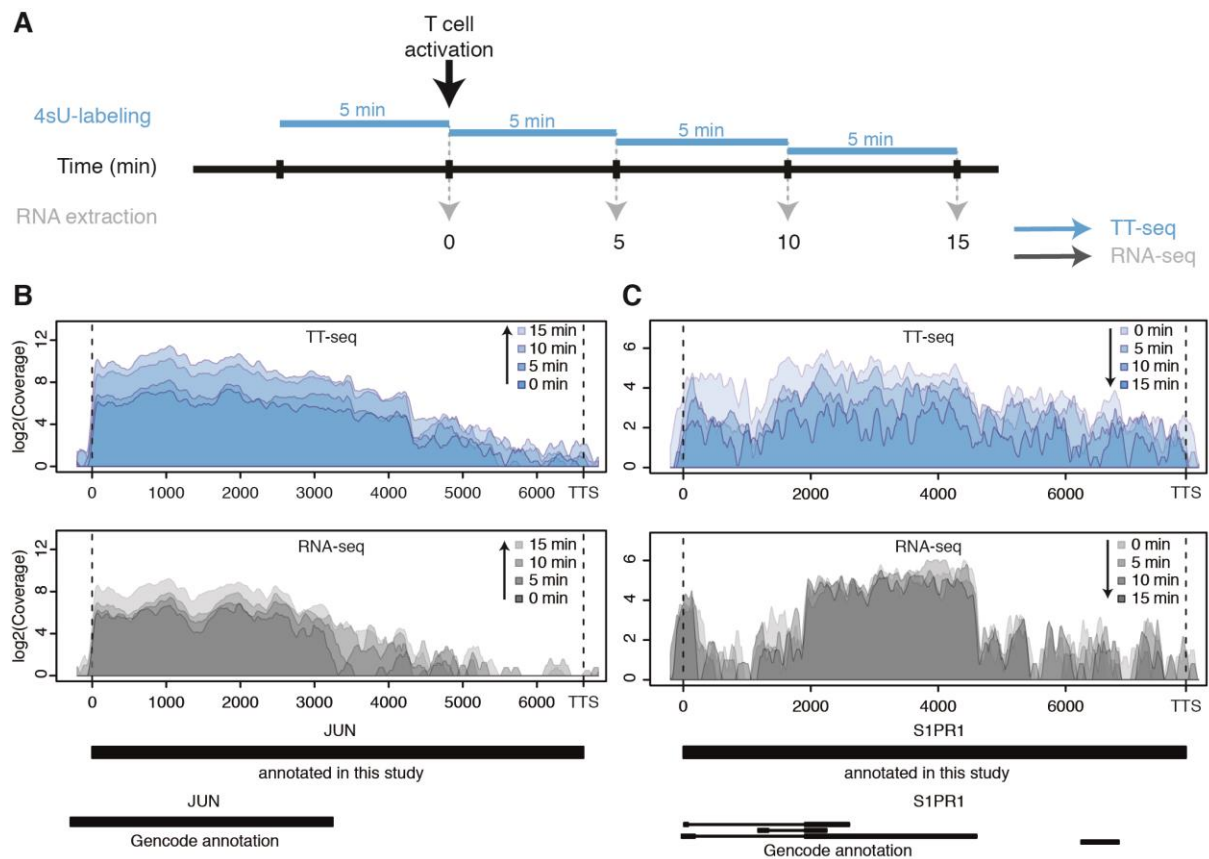


Figure 30: TT-seq during T cell activation. **(A)** Experimental design. **(B)** Genome browser examples for an upregulated mRNA (JUN). Blue coverage: TT-seq data for 0, 5, 10 and 15 min after activation; grey coverage: Total RNA-seq data for 0, 5, 10 and 15 min after activation. **(C)** Same as in **(B)** for a downregulated mRNA (S1PR1).

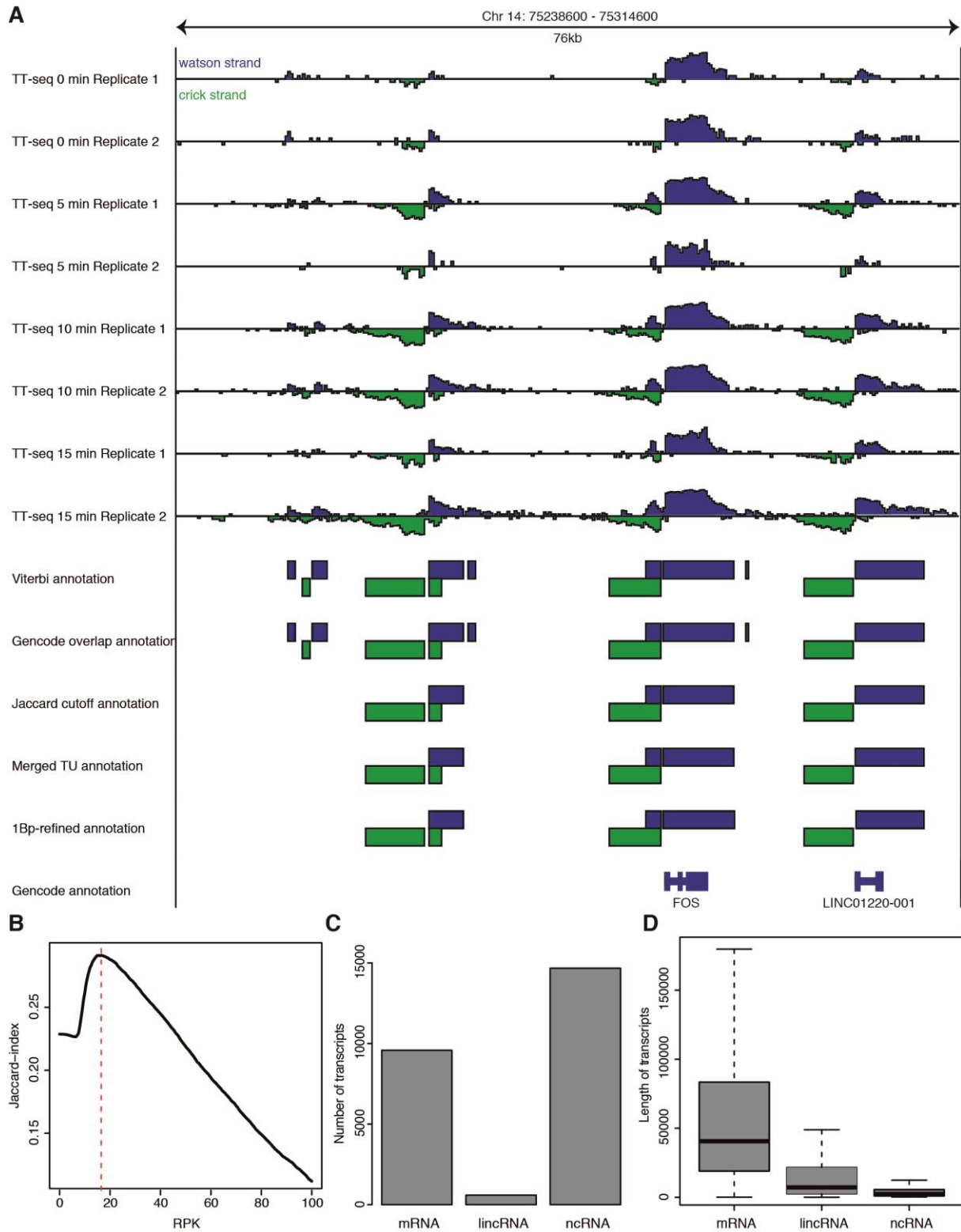


Figure 31: Annotation of transcripts. **(A)** Segmentation workflow. Dark blue: Watson strand, green: crick strand. The top 8 tracks show the antisense corrected TT-seq data tracks (log2 scale), the following tracks indicate the stepwise annotation of transcripts. From the GENCODE annotation, only full transcripts with transcript_support_level 1 are depicted. **(B)** Jaccard index (compared to GENCODE annotation) for different choices of thresholds (x-axis: Reads Per Kilobase (RPK)). **(C)** Number of transcripts per class. **(D)** Distribution of transcript lengths per class.

When performing differential expression analysis, notably not one transcript was significantly up or downregulated compared to time point 0 min after 5, 10 or 15 min in Total RNA-seq (Figure 32A). In strong contrast to this was the analysis of the TT-seq data. We found 276 upregulated and 117 downregulated transcripts after 5 min, 1,811 upregulated and 589 downregulated transcripts after 10 min and 3,412 upregulated and 1,312 downregulated transcripts after 15 min (Figure 32A; $FC > 2$, adjusted $p\text{-value} < 0.05$). Within those differentially expressed transcripts, 337 mRNAs were up and 343 were downregulated after 15 min. The top upregulated mRNAs comprised known markers of T cell activation such as FOS, FOSB, JUN, JUNB and CD69 (Table 17). Other upregulated families included immediate early response genes and other transcription factors such as EGR1, EGR2, EGR3 or NR4A1 (Table 17). Interestingly, a lot of upregulated transcripts were not described in association with T cell activation before, which shows that TT-seq can be used to uncover new target genes in well-studied pathways (Cheadle et al, 2005; Diehn et al, 2002). Enriched GO categories represent responses to stimuli and immune system processes (Figure 32B; MacLeod & Wetzler, 2007). In such a short time period it is expected that general cellular responses are more represented in enriched GO classes rather than specific cellular processes such as T cell activation.

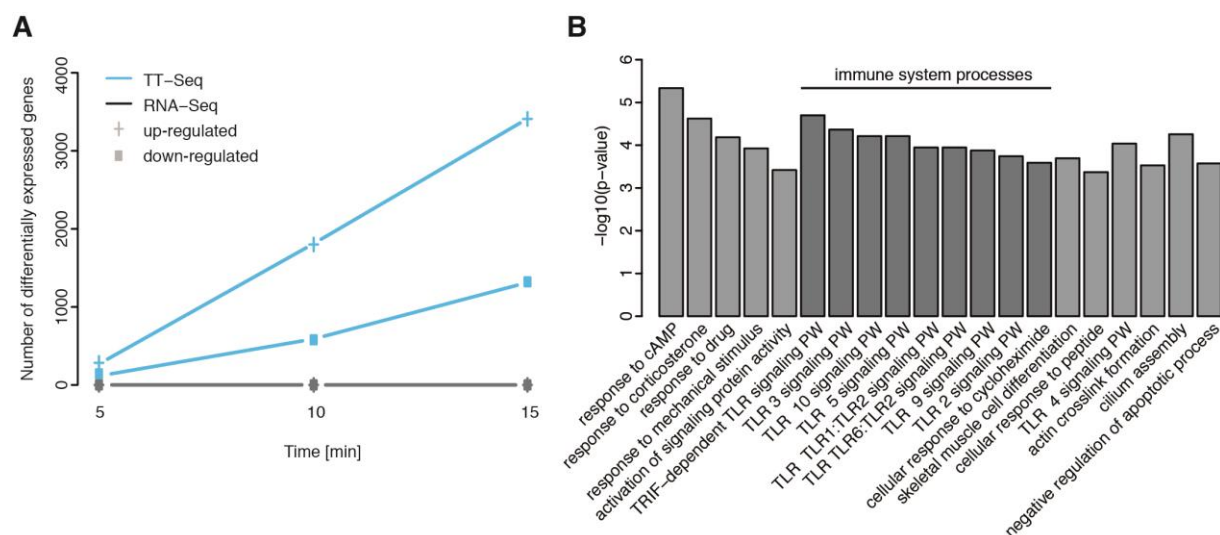


Figure 32: Differentially expressed genes identified by TT-seq. **(A)** Analysis of statistically significant differentially expressed transcripts compared to time point 0 min, before activation. **(B)** GO analysis of upregulated mRNA transcripts between time point 0 and 15 min.

Table 17: Top upregulated protein-coding genes, sorted by their fold change after 15min.

	5 min		10 min		15 min			Diehn, 2002	Cheadle, 2005
Gene Name	log2 FC	adj. P-value	log2 FC	adj. P-value	log2 FC	adj. P-value	ENSG		
FOSB	4.25	8.03E ⁻³⁶	6.60	2.20E ⁻¹²⁵	7.21	9.70E ⁻¹⁰⁷	ENSG00000125740.11	X	X
SERPINE1	1.65	5.67E ⁻⁰⁵	4.17	1.44E ⁻⁵¹	5.88	2.78E ⁻⁷⁵	ENSG00000106366.8		X
GPR50	2.01	2.50E ⁻⁰³	4.11	4.14E ⁻²¹	5.60	7.34E ⁻³³	ENSG00000102195.8		
KLF4	1.94	7.96E ⁻⁰⁸	4.54	5.85E ⁻⁶⁸	5.39	1.04E ⁻⁷⁰	ENSG00000136826.12		
FOS	3.30	1.85E ⁻²⁰	5.20	8.94E ⁻⁷⁵	5.37	3.54E ⁻⁵⁷	ENSG00000170345.7	X	X
IL13RA2	1.67	8.25E ⁻⁰⁵	3.90	5.68E ⁻⁴³	5.08	2.55E ⁻⁵⁶	ENSG00000123496.5		X
CCDC173	1.78	1.00E ⁻⁰⁷	4.15	1.14E ⁻⁶⁷	5.08	1.09E ⁻⁷¹	ENSG00000154479.10		
OSR2	0.6	2.14E ⁻⁰¹	3.14	1.76E ⁻⁴⁹	4.44	5.48E ⁻⁷⁰	ENSG00000164920.7		
TPRG1	2.55	1.82E ⁻⁰⁴	4.00	2.98E ⁻¹⁵	4.30	2.16E ⁻¹⁵	ENSG00000188001.7		
DUSP1	1.4	7.24E ⁻⁰⁵	3.27	4.61E ⁻⁴⁴	4.08	3.17E ⁻⁴⁸	ENSG00000120129.5		
IFNG	1.80	8.14E ⁻⁰³	3.26	5.98E ⁻¹⁴	4.07	8.75E ⁻¹⁸	ENSG00000111537.4	X	
PPP1R15A	1.63	2.30E ⁻⁰⁵	3.14	2.58E ⁻³²	3.93	3.28E ⁻³⁶	ENSG00000087074.7		
EGR2	1.03	3.32E ⁻⁰²	3.00	1.14E ⁻²⁹	3.90	9.51E ⁻³⁶	ENSG00000122877.11	X	X
AC008686.1	1.40	1.76E ⁻⁰¹	3.15	4.25E ⁻⁰⁸	3.90	3.62E ⁻¹¹	ENSG00000280221.1		
EGR3	0.20	7.52E ⁻⁰¹	2.35	2.03E ⁻²⁰	3.70	7.81E ⁻³⁶	ENSG00000179388.8	X	
MASP2	1.60	3.66E ⁻⁰³	2.59	2.91E ⁻¹³	3.65	5.79E ⁻²⁰	ENSG00000009724.14		
LMAN1L	1.21	1.90E ⁻⁰²	2.17	7.92E ⁻¹³	3.62	1.83E ⁻²⁶	ENSG00000140506.14		
NR4A1	0.76	2.71E ⁻⁰¹	2.76	9.95E ⁻¹⁹	3.57	6.99E ⁻²³	ENSG00000123358.17	X	X
JUN	0.87	1.41E ⁻⁰¹	2.38	4.64E ⁻¹⁶	3.44	2.39E ⁻²⁴	ENSG00000177606.6	X	X
RGCC	0.40	4.02E ⁻⁰¹	2.23	1.87E ⁻³⁰	3.43	5.82E ⁻⁵⁰	ENSG00000102760.12		
MTX1	0.81	5.57E ⁻⁰²	2.42	7.45E ⁻²⁸	3.42	5.22E ⁻³⁹	ENSG00000173171.12		
ARC	1.09	8.84E ⁻⁰²	2.76	2.05E ⁻¹⁶	3.25	5.18E ⁻¹⁷	ENSG00000198576.3		
CCDC184	0.96	1.91E ⁻⁰¹	2.66	5.44E ⁻¹⁴	3.14	4.89E ⁻¹⁵	ENSG00000177875.4		
KIAA1683	1.20	6.59E ⁻⁰⁴	2.40	1.49E ⁻²⁶	3.11	2.69E ⁻³¹	ENSG00000130518.14		

eRNA and promoter transcription are correlated

We looked at known enhancers such as the *FOS* locus (Figure 33A) and observed that we can detect enhancers with high accuracy by looking at eRNAs. Furthermore, we also observed that they are differentially expressed during the time course (Figure 33A). This eRNA upregulation achieves up to 350% of the upregulation of the *FOS* transcript during the observed time course (Figure 33A). To be able to classify ncRNAs as eRNAs we made use of the GenoSTAN study (Zacher et al, 2016), where 14 T cell lines were analyzed with regard to their chromatin states. We overlapped all enhancer states with our ncRNAs and categorized 5,572 eRNAs (44% of ncRNAs). Those potential eRNAs were paired with mRNAs in a similar fashion to GREAT (McLean et al, 2010). We then filtered the pairs so to keep the ones within a CTCF-insulated neighborhood (Figure 33B; Hnisz et al, 2016), which lead to 2,553 pairs after filtering (Figure 33C). The transcript expression of pairs within the same CTCF-insulated neighborhood correlated more than for those pairs not within the same neighborhood (Figure 33D).

We analyzed the pairs regarding number of paired transcripts, distance and expression correlation. Altogether, we paired 1,944 eRNAs with 1,446 promoters in 2,553 combinations. On average, 1.7 enhancers regulated a promoter and each enhancer was paired with 1.3 promoters (Figure 34A). The median distance of an eRNA to its paired mRNA is 38 kb (Figure 34B). Expression levels between enhancers and their paired promoters were extremely correlated and the correlation was higher for enhancers located less than 10 kb from their paired promoter (proximal enhancers) rather than further apart (distal enhancers; Figure 34C). We also examined described super-enhancers in Jurkat cells (Hnisz et al, 2013), and found that promoters associated with enhancer that are located within super-enhancers have a higher expression level than other promoters (Figure 34D).

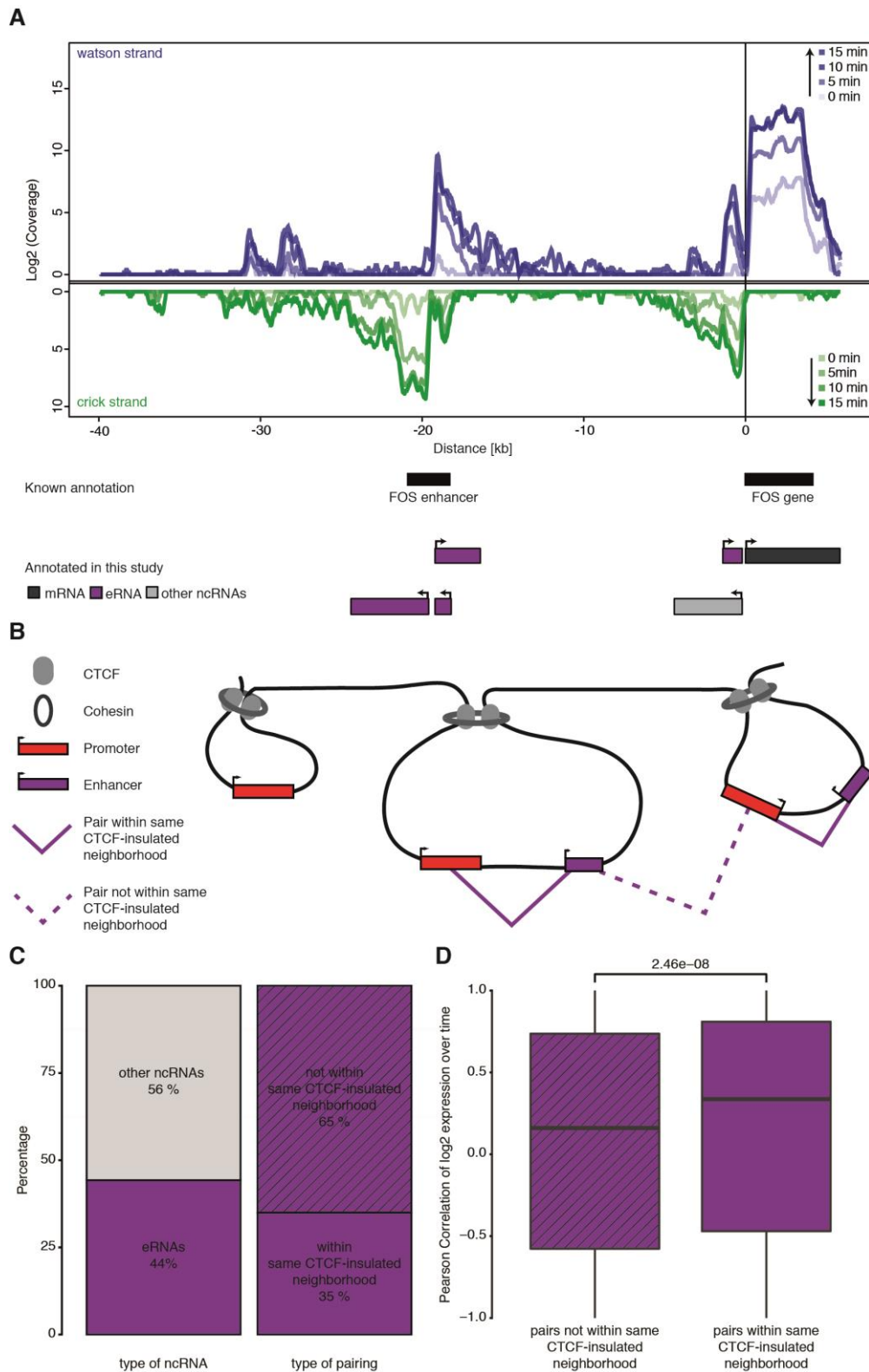


Figure 33: Enhancer identification and pairing with promoters. **(A)** Transcription at FOS and its annotated enhancer. **(B)** Enhancer promoter pairing depending if the pairs are located within one CTCF-insulated neighborhood or not. **(C)** Distribution of identified eRNAs among ncRNAs (left), and of enhancer-promoter pairs within the same CTCF-insulated neighborhood among all pairs. **(D)** Expression correlation between paired enhancers and promoters depending if they are situated in the same insulated neighborhood or not. The p value was derived by two-sided Mann-Whitney U test.

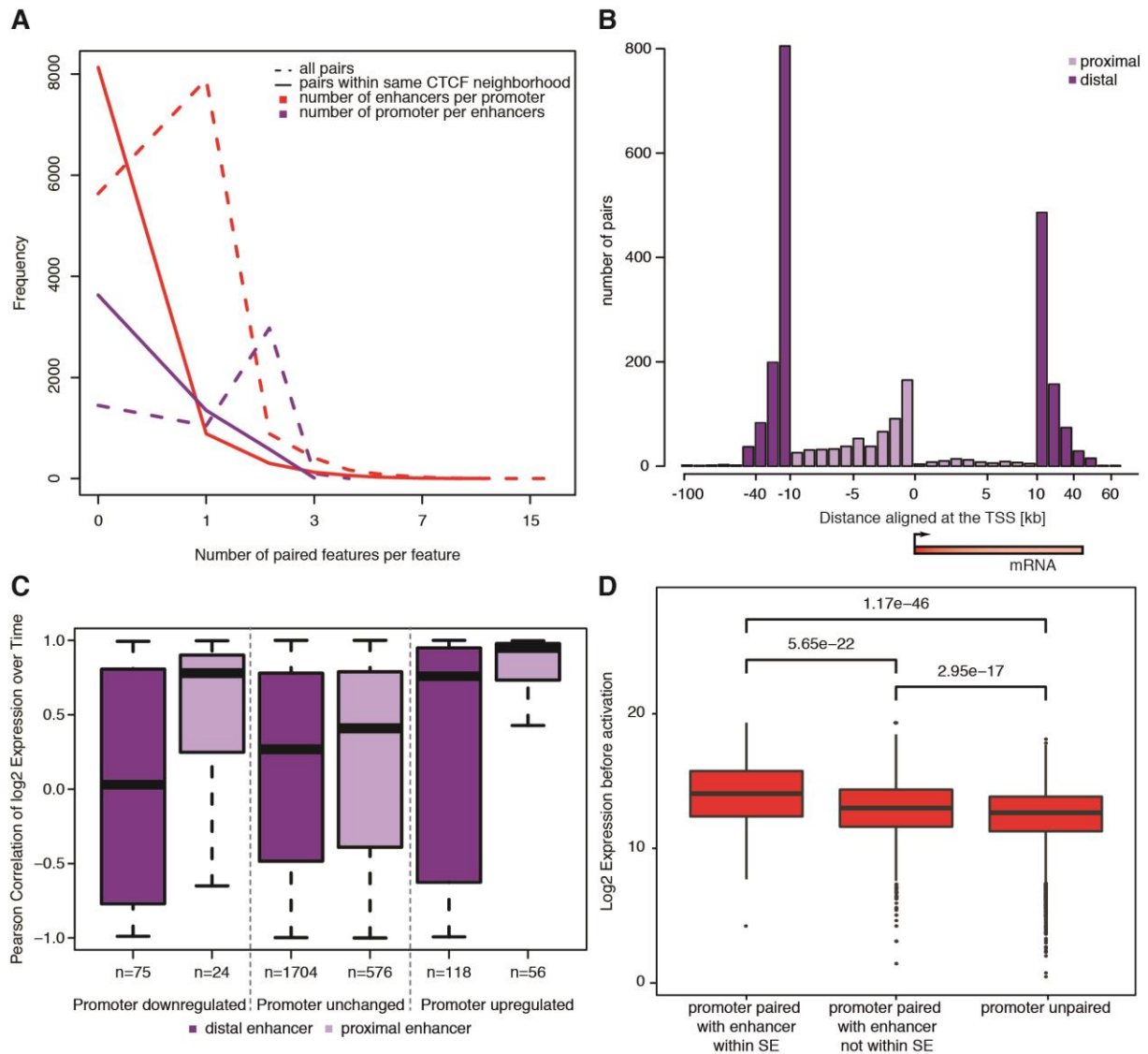


Figure 34: Characteristics of enhancers and promoter pairs. **(A)** Number of enhancer per promoters and promoters per enhancer depending if the pairs are within one CTCF-insulated neighborhood (solid lines) or not (dashed lines). **(B)** Distance distribution between eRNA and mRNA TSS. **(C)** Correlation of expression over time between proximal (left, dark violet) or distal (right, light violet) enhancers and promoters by promoter change (from left to right: downregulated, unchanged, upregulated promoters). **(D)** Expression level of promoters for promoters associated with enhancers within and not within super-enhancers and promoters not associated with any enhancer at all. All p values were derived by two-sided Mann-Whitney U test.

Enhancers and promoters are activated simultaneously

Next, we investigated the temporal activation or shut-down of enhancers and their paired promoters. It has been reported that enhancer transcription precedes the paired promoter's transcription when the second is activated during a time-course (Arner et al, 2015; De Santa et al, 2010b; Kaikkonen et al, 2013). We analyzed our enhancer/promoter pairs after 15 min of activation and observed drastic differences between the Total RNA-seq samples and the TT-

seq samples (Figure 35). Due to high synthesis rate and low half-lives of eRNAs, and high half-lives of mRNAs (Figure 36), the majority of eRNAs precedes their paired mRNAs in the Total RNA-seq samples. This is not the case in the TT-seq samples, where it can be observed that both are either up or downregulated simultaneously (Figure 35). Another evidence that shows that enhancers and their paired promoters are up or downregulated at the same time is seen when we correlated the fold change of the eRNAs and mRNAs at different time points (Figure 37). In the TT-seq samples eRNA and mRNA fold changes correlated the most at the same time points. This is in contrast to the Total RNA-samples, where all mRNAs expression correlated the most with the 5 min eRNA time point, meaning that the fold change of mRNAs reaches only later the same magnitude as the one of the eRNAs.

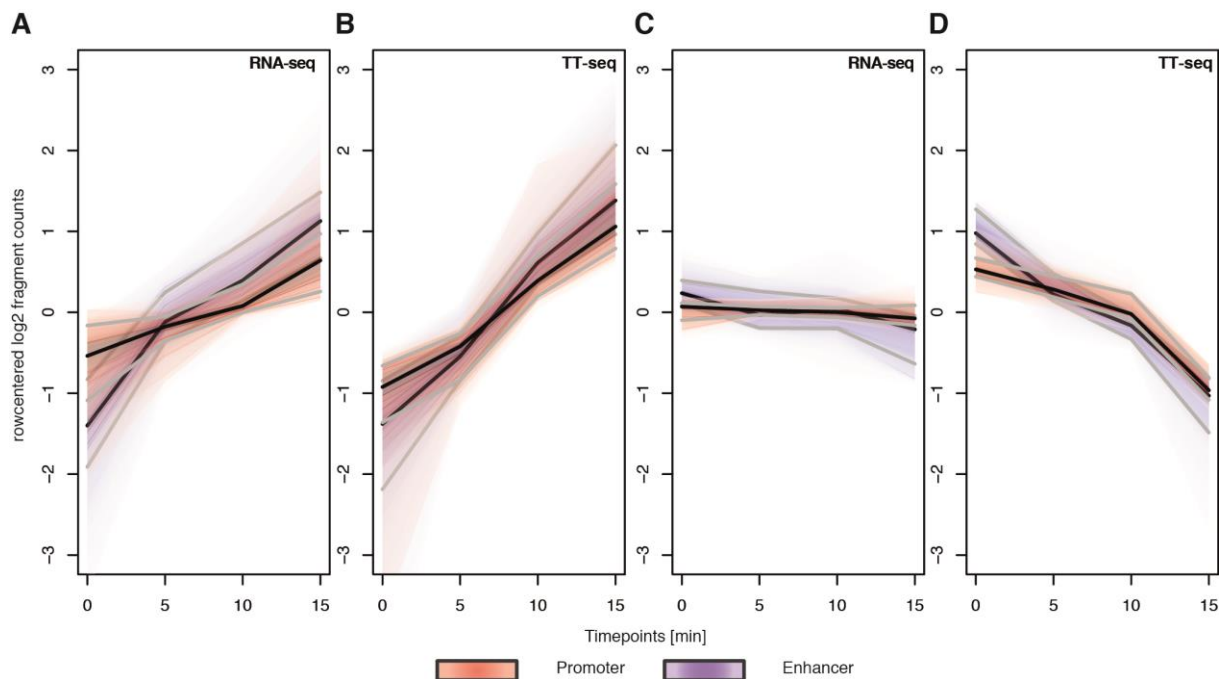


Figure 35: Temporal resolution of enhancer/promoter activation. The black line indicates the median, the lower and upper grey lines the 25 and 75 percent quantiles, respectively. The outermost light shaded lines indicate the 5 and 95 percent quantiles. (A) Promoter/enhancer pairs ($n=71$) that are both upregulated ($FC > 2$) after 15 min activation for RNA-seq samples. (B) Same as in (A) for TT-seq samples. (C) Promoter/enhancer pairs ($n=16$) that are both downregulated ($FC < 2$) after 15 min activation for RNA-seq samples. (D) Same as in (C) for TT-seq samples.

T cell activation responsive enhancers and promoters are poised for activation

CAGE is a very sensitive method and the differences seen in the sequential activation of enhancers and their promoters could be due to the more time-spanning processes that were analyzed (Arner et al, 2015). For late response genes, that are activated hours or days after

stimulation, such as in a cell differentiation process, you first need to open the chromatin environment of the responsible enhancer (Kaikkonen et al, 2013). This enables transcription factor recruitment that will in turn activate the promoter (Spitz & Furlong, 2012). This effect could lead to a time lag between enhancer transcription and promoter transcription. As we only investigate immediate early responses in this study, we propose that the activation process underlies a different mechanism.

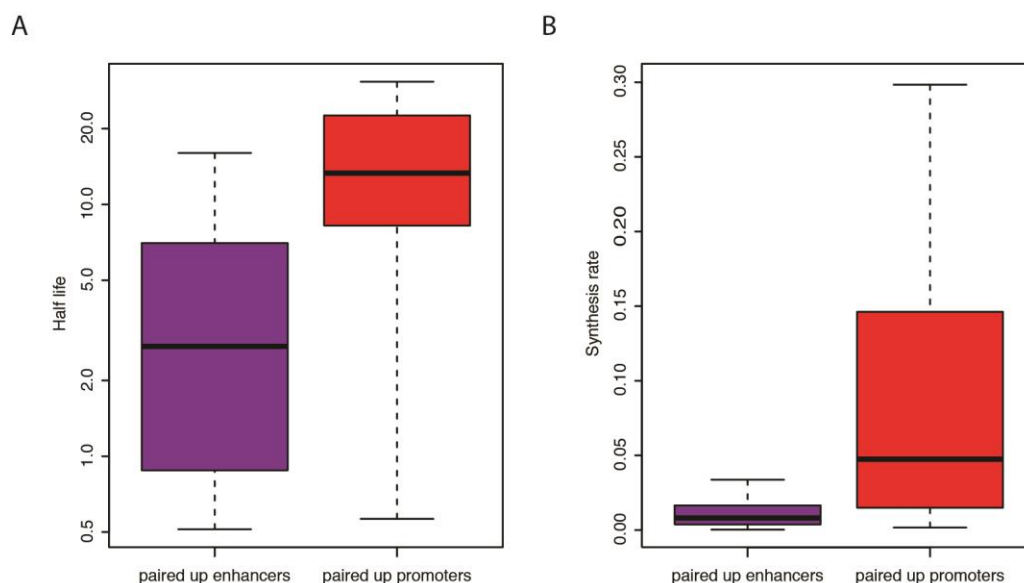


Figure 36: Half-lives and synthesis rates of enhancers and promoters. **(A)** Distribution of half-lives (in minutes) for upregulated enhancers (n=70) and upregulated promoters (n=44) paired in 71 combinations. **(B)** Distribution of synthesis rates (arbitrary scale) for upregulated enhancers (n=70) and upregulated promoters (n=44) paired in 71 combinations.

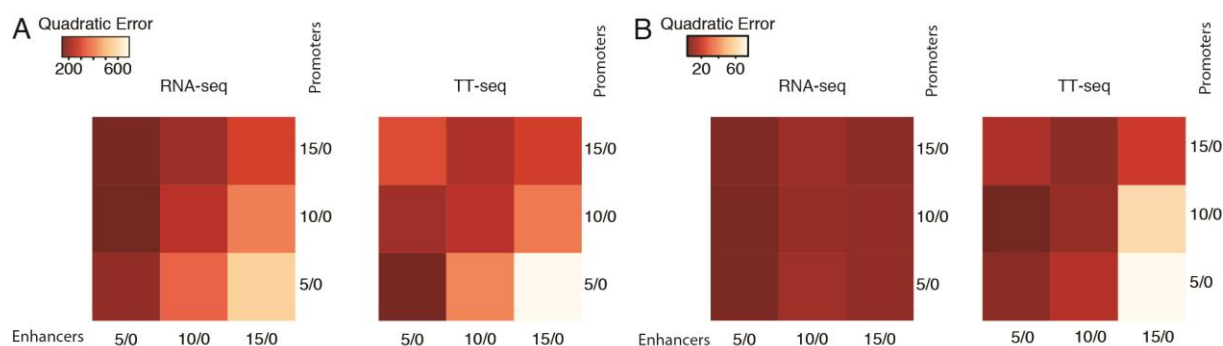


Figure 37: Correlation of fold changes across time between enhancers and promoters. **(A)** Quadratic error correlation between log2 FC for different time points between upregulated Enhancer/Promoter pairs for RNA-seq (left) and TT-seq (right). **(B)** Same as in (A) for downregulated Enhancer/Promoter pairs.

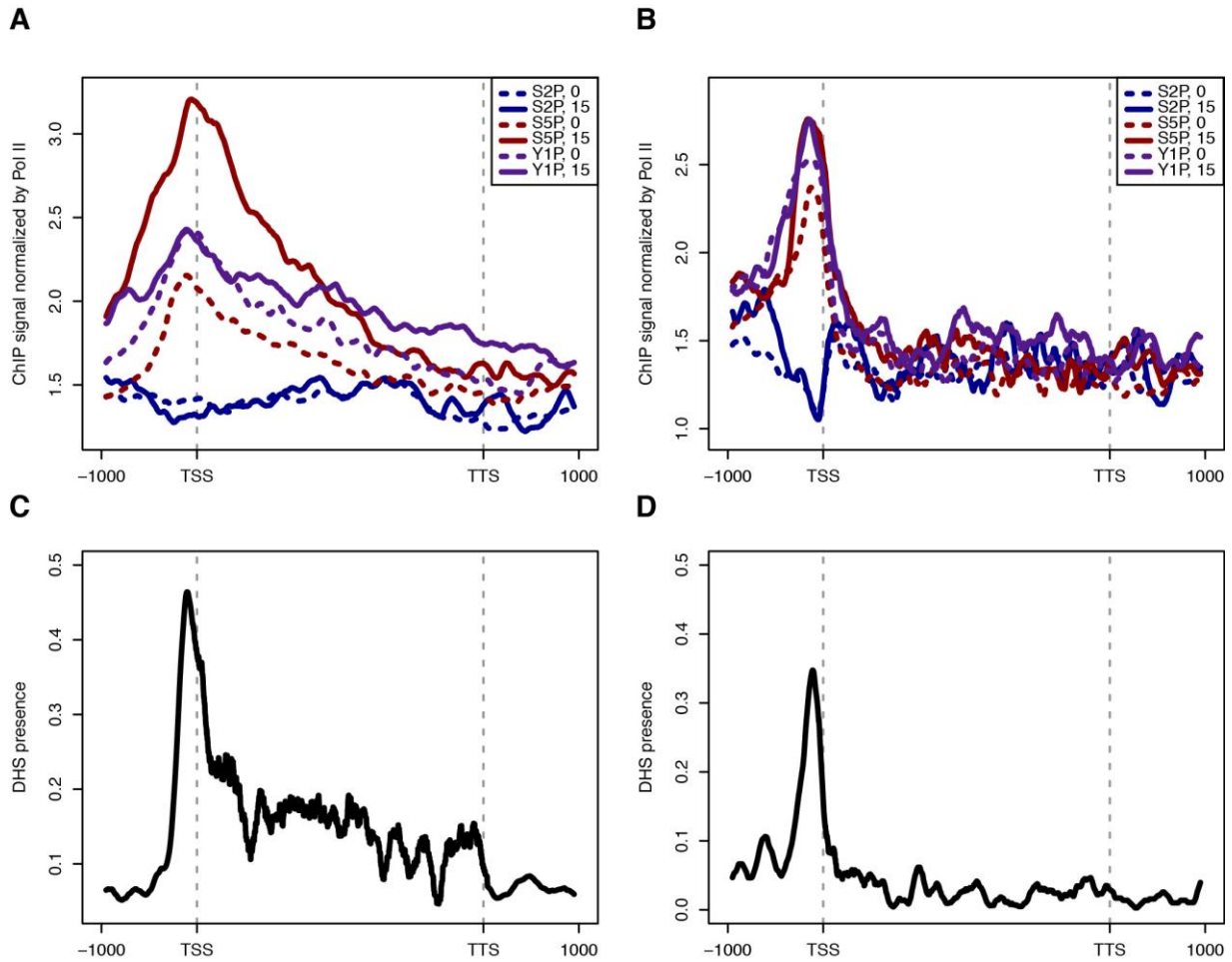


Figure 38: ChIP-seq and DHS analysis at promoters and enhancers that were activated during T cell activation. **(A)** Pol II phospho isoforms occupancy at 311 activated enhancers. Shaded and solid lines represent data collected at time points 0 min and 15 min after activation, respectively. All profiles are normalized by size factor and Pol II density. Blue: S2P, red: S5P, violet: Y1P. **(B)** Same as in (A) for 73 activated promoters. **(C)** Distributions of DHSs at activated enhancers. **(D)** Same as in (C) for activated promoters.

We performed chromatin-immunoprecipitation followed by sequencing (ChIP-seq) for RNA polymerase II (Pol II) and its different phosphorylated forms. We examined enhancers that were not transcribed at the 0 min time point and that were upregulated throughout the time course (lowest 10% of synthesis rates, synthesis rate ≤ 0.005 ; FC ≥ 2). We found that those enhancers are pre-loaded with Pol II that is elongation-competent with its carboxy terminal domain (CTD) phosphorylated on serine 5 (S5P) and tyrosine 1 (Y1P) at time point 0 (Figure 38A). Interestingly, we could not detect phosphorylation on serine 2 (S2P) on activated eRNAs (Figure 38A), similar to observations in yeast on ncRNAs (Kim et al, 2010a). Furthermore, mRNAs that were activated and upregulated throughout the time course were also pre-loaded with elongation-competent Pol II (Figure 38B; synthesis rate ≤ 0.005 and FC ≥ 2). This shows that both enhancers and promoters upregulated in this study are poised for

activation before T cell activation. Additionally, they are already in an open chromatin conformation state as indicated by DNase I hypersensitivity sites (DHSs) present around their TSS (Figure 38C and D) and can rapidly be bound by transcription factors when activated. Together, these results show that the enhancer-promoter landscape in T cells exhibits characteristics that allow for rapid transcription activation.

Discussion

Our results indicate that in a rapid response to environmental changes, such as an immune response, enhancer and promoter transcription happen simultaneously. Due to the sensitivity of the TT-seq protocol in human cells we could observe that enhancers and promoters are activated already after 5 minutes after T cell activation, which shows the need of a method that is capable to detect and quantify transcriptional changes very early after the perturbation of a system.

This study describes a method to globally detect transcripts in a highly dynamic system. TT-seq enables us to detect thousands of differentially expressed transcripts, of which many already have been reported to act in T cell activation. The remaining differentially expressed transcripts are either ncRNAs that we could map de novo, or immediate early genes activated transiently during T cell activation. Our results therefore complement previous genome-wide studies (Cheadle et al, 2005; Diehn et al, 2002), in the same system and help to understand very early transcriptional responses.

Our work yields in the identification and characterization of eRNAs based on their synthesis. Here, we show that the identification and pairing of eRNAs with their respective promoters can be performed by taking advantage of previously published datasets (Hnisz et al, 2016; Zacher et al, 2016), and methodologies (McLean et al, 2010), resulting in highly correlated pairs, regarding their synthesis. In contrast to other genome-wide studies (Arner et al, 2015; De Santa et al, 2010b; Kaikkonen et al, 2013; Schaukowitch et al, 2014), our results indicate a synchronous activation of enhancers and promoters. The temporal resolution of enhancer and promoter transcription is a widely discussed question. Preceding transcription at the enhancer could be argued with the function of the enhancer transcript in releasing Pol II into productive elongation by releasing NELF (Schaukowitch et al, 2014). We, however, propose a model, where both enhancer and promoter are poised for activation. This would lead to a very rapid

transcriptional response, which is relevant in immune system processes (Bahrami & Drablos, 2016). ChIP-seq data support this idea by showing S5P on Pol II at both eRNA and mRNA TSS, indicating a high density of elongation competent Pol II. Additionally, DNase I hypersensitivity sites around eRNA and mRNA TSS indicate that the DNA is already in an open chromatin conformation and thus transcription activation can happen extremely fast when needed. These findings raise the question whether different cellular responses acquired different transcriptional activation strategies.

We expect novel biological insights by the application of TT-seq to other human cell lines, which could help to identify early factors in response signaling pathways, acting in the progression of diseases and cancer or determine cell differentiation.

We demonstrate that TT-seq is suitable for annotating potential eRNAs and quantifying transcriptional changes very early upon stimulation and thus provides insights into gene regulation, activation and enhancer identity. Our data reveals an activation of eRNAs during the T cell activation time course, which correlates with the corresponding promoter activation; thus enhancer and promoter transcription are happening simultaneously.

4 Future Perspectives

The TT-seq methodology finally provides the appropriate tool to investigate transient RNAs in mammals and *in vivo*. Many methods beforehand were sensitive but did not permit analysis of transient RNAs on a genome-wide scale or vice-versa. With the help of TT-seq many open questions such as the ones described below can now be tackled.

4.1 ncRNA surveillance pathway in humans

A vast majority of promoters have been shown to be bidirectional with in one direction producing an mRNA and the other a non-coding RNA (Orehova & Rubtsov, 2013; Xu et al, 2009). Additionally, nucleosome-depleted regions at the 3' end of genes have been shown to harbor promoter features and allow for initiation of transcription (Murray et al, 2012). The genome is therefore largely transcribed, most of which into non-coding transcripts. Regulatory mechanisms such as the Nrd1 surveillance pathway in *Saccharomyces cerevisiae* are responsible for rapid and accurate degradation of non-coding transcripts (Schulz et al, 2013). To this day, the exact degradation pathway for non-coding transcripts in mammals has not been elucidated. It was shown that the exosome pathway can be targeted to some ncRNAs through the NEXT complex (Lubas et al, 2015), or that short transcripts emerging from stalled Pol II are degraded by Xrn2 (Seila et al, 2008). Nevertheless, the degradation pathway of the majority of ncRNAs remains to be discovered. Through knockdown or knockout screening experiments followed by TT-seq, one could measure the amount, emergence or disappearance of ncRNAs in the mutant and as a result learn more about degradation mechanisms.

4.2 Different underlying mechanisms leading to promoter-proximal peaks of Pol II

Promoter-proximal pausing (PPP) is vastly present in higher eukaryotes (Adelman & Lis, 2012). The more it becomes apparent that the majority of metazoan promoters harbors a Pol II ChIP-occupancy peak in their profile, the more the question arises if they all are the result of the same underlying mechanism. For instance, some promoter-proximally paused Pol II do not react to NELF knockdown (Muse et al, 2007). NELF-dependent PPP and the associated ChIP-peak should be abrogated through NELF knockdown. One can therefore speculate that not all Pol II peaks are the result of PPP. Furthermore, it was shown through mathematical

modeling that an accumulation of Pol II near the TSS can be the sheer result of a slow transition from initiation to elongation (Ehrensberger et al, 2013). ChIP on its own is therefore not the appropriate methodology to investigate PPP, since ChIP-peaks are a mixture of two phenomena: amount of crosslinked proteins to a DNA fragment and/or residing time of the same protein on a DNA fragment. TT-seq would help untangle this problem, since one Pol II will produce one RNA fragment regardless of the residing time. One could then differentiate between the amount of Pol II and its residing time. One could also imagine experiments with knockdowns of different PPP players, such as NELF, P-TEFb or DSIF to learn more about PPP. Better than knockdown experiments, which require days when using an RNAi setup, are specific protein inhibitors, which usually act within minutes. One could perform experiments with PTEF-b inhibitors such as DRB or Flavopiridol that more or less specifically target CDK9. This would enable investigation of timing and rate of activation, responsive genes and maybe shed light on other kinases responsible for activation of other gene sets.

4.3 Splicing rate

Splicing has been very difficult to investigate *in vivo* due to the very short half-life of introns. Many methodologies have been used so far: electron microscopy (Beyer & Osheim, 1988), transcription arrest through inhibitors (Clement et al, 1999; Elliott & Rosbash, 1996), *in vitro* reconstituted systems (Padgett et al, 1986), or single-molecule microscopy with reporter splicing assays (Martin et al, 2013). It was shown that cell-free assays have a much slower kinetic than *in vivo* assays (Carmo-Fonseca & Kirchhausen, 2014). On the other hand *in vivo* single-molecule assays are not applicable on a genome-wide scale. TT-seq can solve this problem since it can measure the synthesis and degradation rate at single-nucleotide resolution genome-wide (Schwalb et al, 2016). It has been shown in *S. pombe* that 4sU-seq in combination with spike-ins can be used for determination of splicing rates (Eser et al, 2016). With TT-seq, it will now be possible to determine splicing rates *in vivo* in human cells. One could then analyze the nucleotide composition of splice sites to see if it has an effect on splicing rate, as it has in *S. pombe* (Eser et al, 2016), or perform knockdown experiment of splicing-associated factors to see if they have an impact on rates. Furthermore, one could investigate if different intron lengths lead to different splicing mechanisms or rates.

4.4 Initiation and synthesis rate

In addition to splicing rate, initiation and elongation rates could be elucidated *in vivo* with TT-seq. Studies have made use of metabolic labeling (Neymotin et al, 2014; Schwalb et al, 2012), single-molecule microscopy (Darzacq et al, 2007), inhibitor treatment (Singh & Padgett, 2009) and other methods. Elongation rates have been described to range between 2 and 4 kb/min depending on the methodology used (Ardehali & Lis, 2009). Use of P-TEFb inhibitors such as the CDK9 inhibitors DRB and Flavopiridol could shed light on elongation speed, while CDK7 inhibition through THZ1 could clarify the pausing status of many polymerases (Nilson et al, 2015). Triptolide inhibits XPB of TFIIF, thus hinders bubble opening. In fact, ChIP-seq of human Pol II after triptolide treatment revealed that most genes with paused polymerases are in fact constantly terminating and re-initiating (Chen et al, 2015).

4.5 Pol II CTD modifications at ncRNAs loci

Analysis of transient RNAs under kinase inhibiting conditions would be particularly interesting since the CTD phosphorylation status of Pol II at ncRNAs loci is not entirely clear. Pol II seems to be carrying Ser5P (Gudipati et al, 2008), Ser7P (Kim et al, 2010a) and Tyr1P (Descostes et al, 2014), at most ncRNAs loci. It remains to be investigated if ncRNA loci harbor Pol II phosphorylated at Ser2 or not (Li et al, 2016). It was shown in human cells that termination of uaRNAs is polyA and U1-signal dependent but the Ser2P-status of Pol II was not investigated (Almada et al, 2013; Ntini et al, 2013). Furthermore, it was shown in *S. cerevisiae* that termination and processing of snRNA and snoRNAs are Ser2P-dependent (Egloff et al, 2012; Egloff et al, 2010), but in human, termination mechanism of most ncRNAs remains elusive. Most importantly, the majority of transient RNAs was not known up to this date, or was not known in their full length, rendering Ser2P status investigation unfeasible. In this thesis, we showed that polymerase transcribing eRNAs do not carry Ser2P. It will be interesting to analyze why Ser2P is absent. Is it due to the fact that ncRNAs are relatively short and that Ser2P is usually present far downstream in the gene body, or is it that Ser2P is not required at ncRNAs loci?

4.6 eRNA function

Until now it was difficult to detect eRNAs *in vivo*. This is now possible with our novel TT-seq approach, which not only allows for eRNA detection but additionally enables their full-length mapping (Schwalb et al, 2016). eRNAs are a relatively new RNA class and their

function in cells is unclear (Li et al, 2016). The major question is if eRNA have any function or if they are a mere product of high transcription machineries density at the target promoter, which is spatially linked to the enhancer. Different layers of functions are possible: first, neither the eRNA transcript nor the transcription of the enhancer in itself have any functions. Second, the eRNA transcript has no function but the act of transcribing the enhancer has; since it could lead to a more open chromatin landscape which could increase the strength of the enhancer. Third, the eRNA has a function, for example its sequence composition could lead to the recruitment of proteins that promote transcription of the target promoter or stabilize enhancer-promoter looping. To this day, there is experimental support for all three claims. To shed light on this matter one could make use of the TT-seq technology in combination with the clustered regularly interspaced short palindromic repeats (CRISPR) system, which has made tremendous advances since its description as a tool for molecular biology (Cong et al, 2013; Mali et al, 2013).

The first hypothesis is that neither the eRNA transcript, nor the act of transcribing the eRNA has any function. To this end, one could use the CRISPRi methodology, where the dead effector protein Cas9 (dCas9) is sequence specifically targeted to DNA but has no endonucleolytic cleavage activity and therefore acts as a roadblock for Pol II (Larson et al, 2013; Qi et al, 2013). This leads to a gene downregulation when dCas9 is targeted to the promoter or to the gene body. One could target DNA sites downstream of the TSS of an eRNA; transcription factor binding or initiation on the enhancer would not be hindered but transcription elongation would be blocked and the nascent eRNA would be sequestered to a stalled Pol II (Larson et al, 2013). One could check gene expression with TT-seq; if this has no effect on the target promoter expression, then the first hypothesis holds true.

The second hypothesis is that the eRNA transcript has no function but transcribing its locus has. Since the eRNA has short half-life of 2 min (Schwalb et al, 2016), one would need an experimental setup, where transcription is not hindered but where the eRNAs can be removed, sequestered or degraded within seconds of being produced. Up to this day, experiments have been performed on a small scale and with very long time courses using RNAi. RNAi requires many hours, usually 24h to 48h for it to reach its maximum efficiency. The time scale differences of eRNA half-life and RNAi time requirements seem incompatible to draw valuable biological conclusions. Recently, Cas9 has been modified so that it can be directed to mRNAs (Nelles et al, 2016), and not only to DNA. A cell line expressing RNA-targeting

Cas9 and the according PAMmer could within seconds be directed to an eRNA that is being activated. One would need an eRNA that was off before the experiment and could specifically be activated, for example through T cell activation, EGF stimulation or other rapid cellular responses pathways. Until now, the RNA-targeting Cas9 is only targeted to mRNAs to track it within the cell, but technical advances will without doubt be soon developed to specifically degrade the targeted RNA. With TT-seq, one would analyze the correct down-regulation of the eRNA and check for expression changes of the activated promoter mRNA.

Last hypothesis is that the eRNA transcript in itself carry a function. On the one hand, the very short eRNA half-life and the poor evolutionary conservation argues against the eRNA transcript's functionality (Villar et al, 2015). On the other hand, it has been shown that both the eRNA transcript as a whole (Schaukowitch et al, 2014), or the eRNA's sequence composition in itself (Li et al, 2013), can be important for target mRNA activation. This was only shown on specific loci and was not true for all loci analyzed (Hah et al, 2013). Similar to the study of Li *et al.*, one could scramble the DNA sequence of the eRNA transcript with the help of the CRISPR genome editing technology and see its effect on target gene expression (Li et al, 2013). One could also insert U1 signals within the eRNA, which would improve eRNA stability and see if this has an enhancing effect on target promoter transcription. Last, one could direct RNA-targeting Cas9 to the eRNA or parts of it so that its sequence is masked to proteins. The transcript would still be present, but recruitment of proteins to it would be impaired.

4.7 Concluding remarks

Being able to map transient RNAs *in vivo* in a non-perturbed environment provides a tool to tackle many interesting unanswered questions. These include ncRNA surveillance and degradation pathways, ncRNA transcription cycle and metabolism, eRNA activation, enhancer-promoter specificity, splicing kinetics and many others. Finally, TT-seq will be a very powerful way to try to elucidate eRNA functionality in enhancer-promoter dynamics.

5 References

- Adelman K, Lis JT. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet.* **13**: 720-731.
- Allen BL, Taatjes DJ. (2015). The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol.* **16**: 155-166.
- Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature.* **499**: 360-363.
- Amorim MJ, Cotobal C, Duncan C, Mata J. (2010). Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Mol Syst Biol.* **6**: 380.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, . . . Sandelin A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature.* **507**: 455-461.
- Ardehali MB, Lis JT. (2009). Tracking rates of transcription and splicing in vivo. *Nature structural & molecular biology.* **16**: 1123-1124.
- Armache KJ, Kettenberger H, Cramer P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A.* **100**: 6964-6968.
- Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, . . . Hayashizaki Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science.* **347**: 1010-1014.
- Ashfield R, Patel AJ, Bossone SA, Brown H, Campbell RD, Marcu KB, Proudfoot NJ. (1994). MAZ-dependent termination between closely spaced human complement genes. *EMBO J.* **13**: 5656-5667.
- Bahrami S, Drablos F. (2016). Gene regulation in the immediate-early response process. *Adv Biol Regul*
- Baillat D, Hakimi MA, Naar AM, Shilatifard A, Cooch N, Shiekhattar R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell.* **123**: 265-276.
- Balbo PB, Bohm A. (2007). Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure.* **15**: 1117-1131.
- Banerji J, Rusconi S, Schaffner W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell.* **27**: 299-308.
- Bartkowiak B, Liu P, Phatnani HP, Fuda NJ, Cooper JJ, Price DH, . . . Greenleaf AL. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**: 2303-2316.
- Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, . . . Dreyfuss G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell.* **150**: 53-64.
- Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A.* **99**: 9697-9702.
- Beyer AL, Osheim YN. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* **2**: 754-765.

- Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR. (1999). Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc Natl Acad Sci U S A*. **96**: 7791-7796.
- Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell*. **139**: 547-559.
- Buratowski S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol Cell*. **36**: 541-546.
- Calo E, Wysocka J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell*. **49**: 825-837.
- Calvo O, Manley JL. (2003). Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev*. **17**: 1321-1327.
- Carmo-Fonseca M, Kirchhausen T. (2014). The timing of pre-mRNA splicing visualized in real-time. *Nucleus (Austin, Tex.)*. **5**: 11-14.
- Carneiro T, Carvalho C, Braga J, Rino J, Milligan L, Tollervey D, Carmo-Fonseca M. (2007). Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol Cell Biol*. **27**: 4157-4165.
- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, . . . Gingeras TR. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*. **116**: 499-509.
- Chapman KB, Boeke JD. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell*. **65**: 483-492.
- Cheadle C, Fan J, Cho-Chung YS, Werner T, Ray J, Do L, . . . Becker KG. (2005). Control of gene expression during T cell activation: alternate regulation of mRNA transcription and mRNA stability. *BMC Genomics*. **6**
- Chen F, Gao X, Shilatifard A. (2015). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes Dev*. **29**: 39-47.
- Chen XF, Lehmann L, Lin JJ, Vashisht A, Schmidt R, Ferrari R, . . . Carey M. (2012). Mediator and SAGA have distinct roles in Pol II preinitiation complex assembly and function. *Cell reports*. **2**: 1061-1067.
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, . . . Gingeras TR. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. **308**: 1149-1154.
- Cheung AC, Sainsbury S, Cramer P. (2011). Structural basis of initial RNA polymerase II transcription. *EMBO J*. **30**: 4755-4763.
- Cho EJ, Kobor MS, Kim M, Greenblatt J, Buratowski S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev*. **15**: 3319-3329.

- Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol.* **23**: 232-237.
- Clement JQ, Qian L, Kaplinsky N, Wilkinson MF. (1999). The stability and fate of a spliced intron from vertebrate cells. *RNA (New York, N.Y.).* **5**: 206-220.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, . . . Zhang F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science.* **339**: 819-823.
- Connelly S, Manley JL. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* **2**: 440-452.
- Consortium EP. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489**: 57-74.
- Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A.* **82**: 7934-7938.
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet.* **46**: 1311-1320.
- Core LJ, Waterfall JJ, Lis JT. (2008a). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* **322**: 1845-1848.
- Core LJ, Waterfall JJ, Lis JT. (2008b). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science.* **322**: 1845-1848.
- Costa FF. (2010). Non-coding RNAs: Meet thy masters. *BioEssays : news and reviews in molecular, cellular and developmental biology.* **32**: 599-608.
- Crabtree GR. (1989). Contingent genetic regulatory events in T lymphocyte activation. *Science.* **243**: 355-361.
- Cramer P. (2004). RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev.* **14**: 218-226.
- Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, . . . Vannini A. (2008). Structure of eukaryotic RNA polymerases. *Annual review of biophysics.* **37**: 337-352.
- Crick F. (1970). Central dogma of molecular biology. *Nature.* **227**: 561-563.
- Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. (2007). In vivo dynamics of RNA polymerase II transcription. *Nature structural & molecular biology.* **14**: 796-806.
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, . . . Natoli G. (2010a). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, . . . Natoli G. (2010b). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**: e1000384.
- DeMare LE, Leng J, Cotney J, Reilly SK, Yin J, Sarro R, Noonan JP. (2013). The genomic landscape of cohesin-associated chromatin interactions. *Genome research.* **23**: 1224-1234.

- Descostes N, Heidemann M, Spinelli L, Schuller R, Maqbool MA, Fenouil R, . . . Andrau JC. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife*. **3**: e02105.
- Deshmukh MV, Jones BN, Quang-Dang DU, Flinders J, Floor SN, Kim C, . . . Gross JD. (2008). mRNA decapping is promoted by an RNA-binding channel in Dcp2. *Mol Cell*. **29**: 324-336.
- Devaiah BN, Lewis BA, Cherman N, Hewitt MC, Albrecht BK, Robey PG, . . . Singer DS. (2012). BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A*. **109**: 6927-6932.
- Dhanasekaran K, Kumari S, Kanduri C. (2013). Noncoding RNAs in chromatin organization and transcription regulation: an epigenetic view. *Sub-cellular biochemistry*. **61**: 343-372.
- Diehn M, Alizadeh AA, Rando OJ, Liu CL, Stankunas K, Botstein D, . . . Brown PO. (2002). Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation. *Proc Natl Acad Sci U S A*. **99**: 11796-11801.
- Diribarne G, Bensaude O. (2009). 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA biology*. **6**: 122-128.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, . . . Gingeras TR. (2012). Landscape of transcription in human cells. *Nature*. **489**: 101-108.
- Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, . . . Koszinowski UH. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA (New York, N.Y.)*. **14**: 1959-1972.
- Doma MK, Parker R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. **440**: 561-564.
- Eddy SR. (2001). Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*. **2**: 919-929.
- Egloff S, Dienstbier M, Murphy S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends in Genetics*
- Egloff S, Murphy S. (2008). Cracking the RNA polymerase II CTD code. *Trends in genetics : TIG*. **24**: 280-288.
- Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, . . . Murphy S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*. **318**: 1777-1779.
- Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S. (2010). The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem*. **285**: 20564-20569.
- Ehrensberger AH, Kelly GP, Svejstrup JQ. (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell*. **154**: 713-715.
- Eick D, Geyer M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews*. **113**: 8456-8490.
- Elliott DJ, Rosbash M. (1996). Yeast pre-mRNA is composed of two populations with distinct kinetic properties. *Exp Cell Res*. **229**: 181-188.

- Ellisen LW, Palmer RE, Maki RG, Truong VB, Tamayo P, Oliner JD, Haber DA. (2001). Cascades of transcriptional induction during human lymphocyte activation. *European Journal of Cell Biology*. **80**: 321-328.
- Eser P, Demel C, Maier KC, Schwalb B, Pirkel N, Martin DE, . . . Tresch A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Molecular Systems Biology*. **10**: 717.
- Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, . . . Gagneur J. (2016). Determinants of RNA metabolism in the *Schizosaccharomyces pombe* genome. *Mol Syst Biol*. **12**: 857.
- Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA (New York, N.Y.)*. **15**: 21-32.
- Fan X, Chou DM, Struhl K. (2006). Activator-specific recruitment of Mediator in vivo. *Nature structural & molecular biology*. **13**: 117-120.
- Feske S, Giltzane J, Dolmetsch R, Staudt LM, Rao A. (2001). Gene regulation mediated by calcium signals in T lymphocytes. *Nat Immunol*. **2**: 316-324.
- Flynn RA, Almada AE, Zamudio JR, Sharp PA. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences of the United States of America*. **108**: 10460-10465.
- Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, . . . Bentley DL. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell*. **60**: 256-267.
- Garcia-Martinez J, Aranda A, Perez-Ortin JE. (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell*. **15**: 303-313.
- Gil A, Proudfoot NJ. (1984). A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. *Nature*. **312**: 473-474.
- Gilchrist DA, Nechaev S, Lee C, Ghosh SK, Collins JB, Li L, . . . Adelman K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev*. **22**: 1921-1933.
- Gilmour DS, Lis JT. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol Cell Biol*. **6**: 3984-3989.
- Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol*. **24**: 5534-5547.
- Gu B, Eick D, Bensaude O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Res*. **41**: 1591-1603.

- Gudipati RK, Villa T, Boulay J, Libri D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology*. **15**: 786-794.
- Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, . . . Merckenschlager M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. **460**: 410-413.
- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome research*. **23**: 1210-1223.
- Hahn S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology*. **11**: 394-403.
- Halbach F, Reichelt P, Rode M, Conti E. (2013). The yeast ski complex: crystal structure and RNA channeling to the exosome complex. *Cell*. **154**: 814-826.
- Hargrove JL, Schmidt FH. (1989). The role of mRNA and protein stability in gene expression. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **3**: 2360-2370.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, . . . Hubbard TJ. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research*. **22**: 1760-1774.
- Heidemann M, Hintermair C, Voss K, Eick D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta*. **1829**: 55-62.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, . . . Ren B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. **39**: 311-318.
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*. **308**: 118-120.
- Herrick D, Parker R, Jacobson A. (1990). Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol*. **10**: 2269-2284.
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, . . . Young RA. (2013). Super-enhancers in the control of cell identity and disease. *Cell*. **155**: 934-947.
- Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, . . . Young RA. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. **351**: 1454-1458.
- Holstege FC, Fiedler U, Timmers HT. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J*. **16**: 7468-7480.
- Houseley J, LaCava J, Tollervey D. (2006). RNA-quality control by the exosome. *Nat Rev Mol Cell Biol*. **7**: 529-539.
- Houseley J, Tollervey D. (2009). The many pathways of RNA degradation. *Cell*. **136**: 763-776.
- Hsieh CL, Fei T, Chen Y, Li T, Gao Y, Wang X, . . . Kantoff PW. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci U S A*. **111**: 7319-7324.
- Isken O, Maquat LE. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet*. **9**: 699-712.

- Jack HM, Wabl M. (1988). Immunoglobulin mRNA stability varies during B lymphocyte differentiation. *EMBO J.* **7**: 1041-1046.
- Jensen TH, Jacquier A, Libri D. (2013). Dealing with pervasive transcription. *Mol Cell.* **52**: 473-484.
- Johnson DS, Mortazavi A, Myers RM, Wold B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science.* **316**: 1497-1502.
- Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature.* **468**: 664-668.
- Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, . . . Glass CK. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell.* **51**: 310-325.
- Keene RG, Mueller A, Landick R, London L. (1999). Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N- and c-myc genes. *Nucleic Acids Res.* **27**: 3173-3182.
- Kettenberger H, Armache KJ, Cramer P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell.* **16**: 955-965.
- Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, . . . Bentley DL. (2010a). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology.* **17**: 1279-1286.
- Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedeia E, Greenblatt JF, Buratowski S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature.* **432**: 517-522.
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, . . . Greenberg ME. (2010b). Widespread transcription at neuronal activity-regulated enhancers. *Nature.* **465**: 182-187.
- Kireeva ML, Komissarova N, Waugh DS, Kashlev M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem.* **275**: 6530-6536.
- Klauer AA, van Hoof A. (2012). Degradation of mRNAs that lack a stop codon: a decade of nonstop progress. *Wiley Interdiscip Rev RNA.* **3**: 649-660.
- Knowling S, Morris KV. (2011). Non-coding RNA and antisense RNA. Nature's trash or treasure? *Biochimie.* **93**: 1922-1927.
- Kodius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, . . . Carninci P. (2006). CAGE: cap analysis of gene expression. *Nat Methods.* **3**: 211-222.
- Komarnitsky P, Cho EJ, Buratowski S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14**: 2452-2460.
- Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell.* **14**: 387-394.
- Kuehner JN, Pearson EL, Moore C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol.* **12**: 283-294.

- Kwak H, Fuda NJ, Core LJ, Lis JT. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*. **339**: 950-953.
- Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. **8**: 2180-2196.
- Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, . . . Rosenfeld MG. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*. **498**: 516-520.
- Li W, Notani D, Rosenfeld MG. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet*. **17**: 207-223.
- Lian Z, Karpikov A, Lian J, Mahajan MC, Hartman S, Gerstein M, . . . Weissman SM. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome research*. **18**: 1224-1237.
- Lidschreiber M, Leike K, Cramer P. (2013). Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Molecular and Cellular Biology*. **33**: 3805-3816.
- Liu Y, Kung C, Fishburn J, Ansari AZ, Shokat KM, Hahn S. (2004). Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol Cell Biol*. **24**: 1721-1735.
- Logan J, Falck-Pedersen E, Darnell JE, Jr., Shenk T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A*. **84**: 8306-8310.
- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell reports*. **10**: 178-192.
- Luse DS. (2013). Promoter clearance by RNA polymerase II. *Biochim Biophys Acta*. **1829**: 63-68.
- MacLeod H, Wetzler LM. (2007). T cell activation by TLRs: a role for TLRs in the adaptive immune response. *Sci STKE*. **2007**: pe48.
- Makino DL, Schuch B, Stegmann E, Baumgartner M, Basquin C, Conti E. (2015). RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature*. **524**: 54-58.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, . . . Church GM. (2013). RNA-guided human genome engineering via Cas9. *Science*. **339**: 823-826.
- Malik S, Roeder RG. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci*. **30**: 256-263.
- Maquat LE, Tarn WY, Isken O. (2010). The pioneer round of translation: features and functions. *Cell*. **142**: 368-374.
- Marrack P, Mitchell T, Hildeman D, Kedl R, Teague TK, Bender J, . . . Kappler J. (2000). Genomic-scale analysis of gene expression in resting and activated T cells. *Curr Opin Immunol*. **12**: 206-209.

- Martin RM, Rino J, Carvalho C, Kirchhausen T, Carmo-Fonseca M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell reports*. **4**: 1144-1155.
- Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J*. **30**: 1302-1310.
- Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, . . . Churchman LS. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*. **161**: 541-554.
- Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nature structural & molecular biology*. **17**: 1272-1278.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, . . . Bejerano G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. **28**: 495-501.
- Melgar MF, Collins FS, Sethupathy P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol*. **12**: R113.
- Michel M, Demel C, Hofmann K, Sawicka A, Zacher B, Schwalb B, . . . Cramer P. (2016). TT-seq captures simultaneous activation of eRNAs and promoters during T cell activation. *Manuscript under preparation*.
- Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, . . . Cramer P. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol*. **7**: 458.
- Mischo HE, Proudfoot NJ. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta*. **1829**: 174-185.
- Murray SC, Serra Barros A, Brown DA, Dudek P, Ayling J, Mellor J. (2012). A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res*. **40**: 2432-2444.
- Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, . . . Adelman K. (2007). RNA polymerase is poised for activation across the genome. *Nat Genet*. **39**: 1507-1511.
- Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell*. **19**: 3418-3436.
- Navarro P, Avner P. (2010). An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells. *BioEssays : news and reviews in molecular, cellular and developmental biology*. **32**: 581-588.
- Nechaev S, Adelman K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta*. **1809**: 34-45.
- Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, Yeo GW. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*. **165**: 488-496.

- Neymotin B, Athanasiadou R, Gresham D. (2014). Determination of in vivo RNA kinetics using RATE-seq. *RNA (New York, N.Y.)*. **20**: 1645-1652.
- Nilson KA, Guo J, Turek ME, Brogie JE, Delaney E, Luse DS, Price DH. (2015). THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol Cell*. **59**: 576-587.
- Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, . . . Proudfoot NJ. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*. **161**: 526-540.
- Nonet M, Scafe C, Sexton J, Young R. (1987). Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol*. **7**: 1602-1611.
- Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, . . . Jensen TH. (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nature structural & molecular biology*. **20**: 923-928.
- Ong CT, Corces VG. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. **15**: 234-246.
- Orekhova AS, Rubtsov PM. (2013). Bidirectional promoters in the transcription of mammalian genomes. *Biochemistry. Biokhimiia*. **78**: 335-341.
- Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA. (1986). Splicing of messenger RNA precursors. *Annu Rev Biochem*. **55**: 1119-1150.
- Palazzo AF, Lee ES. (2015). Non-coding RNA: what is functional and what is junk? *Front Genet*. **6**: 2.
- Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, . . . Merkenschlager M. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*. **132**: 422-433.
- Perez-Ortin JE. (2007). Genomics of mRNA turnover. *Brief Funct Genomic Proteomic*. **6**: 282-291.
- Perez-Ortin JE, Alepuz PM, Moreno J. (2007). Genomics and gene transcription kinetics in yeast. *Trends in genetics : TIG*. **23**: 250-257.
- Peterlin BM, Price DH. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*. **23**: 297-305.
- Petrascheck M, Escher D, Mahmoudi T, Verrijzer CP, Schaffner W, Barberis A. (2005). DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res*. **33**: 3743-3750.
- Plaschka C, Lariviere L, Wenzek L, Seizl M, Hemann M, Tegunov D, . . . Cramer P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*. **518**: 376-380.
- Porrua O, Libri D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol*. **16**: 190-202.
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, . . . Jensen TH. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science*. **322**: 1851-1854.
- Proudfoot NJ. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci*. **14**: 105-110.

- Proudfoot NJ. (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* **25**: 1770-1782.
- Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **152**: 1173-1183.
- Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, . . . Regev A. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell.* **159**: 1698-1710.
- Raghavan A, Ogilvie RL, Reilly C, Abelson ML, Raghavan S, Vasdevani J, . . . Bohjanen PR. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Research.* **30**: 5529-5538.
- Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, . . . Pikaard CS. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell.* **33**: 192-203.
- Ringel R, Sologub M, Morozov YI, Litonin D, Cramer P, Temiakov D. (2011). Structure of human mitochondrial RNA polymerase. *Nature.* **478**: 269-273.
- Roeder RG, Rutter WJ. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature.* **224**: 234-237.
- Rogge L, Bianchi E, Biffi M, Bono E, Chang SY, Alexander H, . . . Certa U. (2000). Transcript imaging of the development of human T helper cells using oligonucleotide arrays. *Nat Genet.* **25**: 96-101.
- Ross J. (1995). mRNA stability in mammalian cells. *Microbiol Rev.* **59**: 423-450.
- Rougemaille M, Villa T, Gudipati RK, Libri D. (2008). mRNA journey to the cytoplasm: attire required. *Biol Cell.* **100**: 327-342.
- Sainsbury S, Bernecky C, Cramer P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol.* **16**: 129-143.
- Sarge KD, Park-Sarge OK. (2005). Gene bookmarking: keeping the pages open. *Trends Biochem Sci.* **30**: 605-610.
- Schaeffer D, van Hoof A. (2011). Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proc Natl Acad Sci U S A.* **108**: 2366-2371.
- Schaukowitch K, Joo JY, Liu X, Watts JK, Martinez C, Kim TK. (2014). Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol Cell.* **56**: 29-42.
- Schmid M, Jensen TH. (2013). Transcription-associated quality control of mRNP. *Biochim Biophys Acta.* **1829**: 158-168.
- Schubeler D. (2007). Enhancing genome annotation with chromatin. *Nat Genet.* **39**: 284-285.
- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, . . . Cramer P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell.* **155**: 1075-1087.
- Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, . . . Cramer P. (2016). TT-seq maps the human transient transcriptome. *Science.* **352**: 1225-1228.

- Schwalb B, Schulz D, Sun M, Zacher B, Dumcke S, Martin DE, . . . Tresch A. (2012). Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA). *Bioinformatics*. **28**: 884-885.
- Schwede A, Ellis L, Luther J, Carrington M, Stoecklin G, Clayton C. (2008). A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. *Nucleic Acids Res*. **36**: 3374-3388.
- Schwer B, Mao X, Shuman S. (1998). Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. *Nucleic Acids Res*. **26**: 2050-2057.
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, . . . Sharp PA. (2008). Divergent transcription from active promoters. *Science*. **322**: 1849-1851.
- Shandilya J, Roberts SG. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta*. **1819**: 391-400.
- Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, . . . Hayashizaki Y. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*. **100**: 15776-15781.
- Shlyueva D, Stampfel G, Stark A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. **15**: 272-286.
- Singh J, Padgett RA. (2009). Rates of in situ transcription and splicing in large human genes. *Nature structural & molecular biology*. **16**: 1128-1133.
- Skaar JR, Ferris AL, Wu X, Saraf A, Khanna KK, Florens L, . . . Pagano M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. *Cell research*. **25**: 288-305.
- Sonenberg N, Hinnebusch AG. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*. **136**: 731-745.
- Spitz F, Furlong EE. (2012). Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. **13**: 613-626.
- Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, . . . de Laat W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. **20**: 2349-2354.
- Stalder L, Muhlemann O. (2008). The meaning of nonsense. *Trends Cell Biol*. **18**: 315-321.
- Steinmetz EJ, Brow DA. (1996). Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol*. **16**: 6993-7003.
- Sun M, Schwalb B, Pirkel N, Maier KC, Schenk A, Failmezger H, . . . Cramer P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell*. **52**: 52-62.
- Sun M, Schwalb B, Schulz D, Pirkel N, Etzold S, Lariviere L, . . . Cramer P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome research*. **22**: 1350-1359.
- Svejstrup JQ, Li Y, Fellows J, Gnatt A, Bjorklund S, Kornberg RD. (1997). Evidence for a mediator cycle at the initiation of transcription. *Proc Natl Acad Sci U S A*. **94**: 6075-6078.

- Tange TO, Nott A, Moore MJ. (2004). The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol.* **16**: 279-284.
- Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell.* **23**: 853-864.
- Thomas MC, Chiang CM. (2006). The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology.* **41**: 105-178.
- Vadie N, Saayman S, Lenox A, Ackley A, Clemson M, Burdach J, . . . Morris KV. (2015). MYCNOS functions as an antisense RNA regulating MYCN. *RNA biology.* **12**: 893-899.
- Valen E, Preker P, Andersen PR, Zhao X, Chen Y, Ender C, . . . Jensen TH. (2011). Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nature structural & molecular biology.* **18**: 1075-1082.
- Vasiljeva L, Buratowski S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell.* **21**: 239-248.
- Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology.* **15**: 795-804.
- Venters BJ, Pugh BF. (2013). Genomic organization of human transcription initiation complexes. *Nature.* **502**: 53-58.
- Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, . . . Odom DT. (2015). Enhancer evolution across 20 mammalian species. *Cell.* **160**: 554-566.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, . . . Pennacchio LA. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* **457**: 854-858.
- Wang W, Carey M, Gralla JD. (1992). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. *Science.* **255**: 450-453.
- Wang Y, Fairley JA, Roberts SG. (2010). Phosphorylation of TFIIB links transcription initiation and termination. *Current biology : CB.* **20**: 548-553.
- Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. (2002). Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A.* **99**: 5860-5865.
- Wang Z, Gerstein M, Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* **10**: 57-63.
- Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF. (2005). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol.* **23**: 1383-1390.
- Wei C, Moss B. (1977). 5'-Terminal capping of RNA by guanylyltransferase from HeLa cell nuclei. *Proc Natl Acad Sci U S A.* **74**: 3758-3761.
- Wei CM, Gershowitz A, Moss B. (1975). Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell.* **4**: 379-386.
- West S, Gromak N, Proudfoot NJ. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature.* **432**: 522-525.

- West S, Proudfoot NJ, Dye MJ. (2008). Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Molecular Cell*. **29**: 600-610.
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, . . . Bahler J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*. **453**: 1239-1243.
- Wu H, Nord AS, Akiyama JA, Shoukry M, Afzal V, Rubin EM, . . . Visel A. (2014). Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS Genet*. **10**: e1004610.
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, . . . Steinmetz LM. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*. **457**: 1033-1037.
- Zacher B, Michel M, Schwalb B, Cramer P, Tresch A, Gagneur J. (2016). Accurate promoter and enhancer identification in 127 ENCODE and Roadmap Epigenomics cell types and tissues by GenoSTAN. *bioRxiv*
- Zeidan Q, Hart GW. (2010). The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *Journal of cell science*. **123**: 13-22.
- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, . . . Young RA. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet*. **39**: 1512-1516.
- Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *J Biol Chem*. **287**: 8541-8551.
- Zylber EA, Penman S. (1971). Products of RNA polymerases in HeLa cell nuclei. *Proc Natl Acad Sci U S A*. **68**: 2861-2865.

6 Abbreviations

4sU	4-thiouridine
Amp	Ampicillin
bp	base pairs
Cdk	cyclin-dependent kinase
ChIP-seq	Chromatin immune-precipitation followed by deep-sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
eRNA	Enhancer RNA
FBS	Fetal bovine serum
fwd	Forward primer
kb	kilobase
lncRNAs	long non-coding RNA
nt	nucleotide
pA	Polyadenylation
Pen-Strep	Penicillin-streptomycin
PMA	Phorbol 12-myristate 13-acetate
Pol	Polymerase
PPP	Promoter-proximal pausing
rev	Reverse primer
RT	Room-temperature
Ser2P, Ser5P, Ser7P	Serine 2, Serine 5 and Serine 7 phosphorylation
TADs	Topologically associated domains
TSS	Transcription start site
TT-seq	Transient-transcriptome sequencing
Tyr1P	Tyrosine 1 phosphorylation
uaRNAs	Upstream antisense RNAs
v/v %	volume/volume percent
w/v %	weight/volume percent

Acknowledgments

I am very grateful to Patrick Cramer for giving me the opportunity to work in his lab. Foremost, for believing in my capabilities and making me rise to the challenge of establishing the human systems biology in his lab. His trust, advice and the working atmosphere he created were crucial to the abilities I have acquired. His mentoring was the perfect balance between support and independence, which creates an ideal space for personal growth. Last, his talent to choose employees makes his lab, not only very mentally stimulating, but a truly great atmosphere to work in. I have encountered in this lab so many kind-hearted, spirited and intelligent persons over this past 5 years; it was truly a pleasure to come to work.

I would like to thank Julien Gagneur, whose great instinct and idea led to the development of TT-seq. His scrutinous eye was key during the discovery phase of the analyses and his many advices and ideas were very valuable to my projects.

I also would like to thank other members of my examination board: PD Dr. Dietmar Martin, Prof. Dr. Peter Becker, Prof. Dr. Karl-Peter Hopfner and Prof. Dr. Klaus Förstemann, for their advice and time.

My deepest gratitude goes to Stefan Krebs from the sequencing facility at the Gene Center, where I was a daily guest during my experimental peak time. He processed hundreds of samples, cheered me up when it didn't work and has the greatest expert knowledge I have ever accoutered in sequencing and sequencing troubleshooting. Thank you!

Björn, Bene and Carina were amazing to work with. I am grateful for these team works, for all the help, the discussions and the crazy theories. Not to forget the great taste in interior design and the many compliments it got our office.

I also would like to thank Katja, Kerstin and Steffi E. – my 4sU support team. I think the four of us tested every μl , μg and $^{\circ}\text{C}$ of this protocol in Munich. And the story goes on in Göttingen and Stockholm! I was really lucky to have them and to have benefited from their experience. On top, Katja helped so much with advice, spike-ins and materials when ours were empty and Kerstin was the best bench and desk neighbor for my 3 Munich years.

I also thank the Cramer lab as a whole, with all of his former and current lab members and the extraordinary atmosphere and team spirit that I hope will subsist as the lab grows larger. I especially thank Wolfgang, Michi, Björn, Katja, Merle, Jürgen, Sofia, Carina and Saskia for accompanying me through my PhD and all the great times outside of the lab.

I thank Claudia, Janine and Kirsten for taking care of the lab and its administration.

I am very grateful for my friends that always make everything better, easier and funnier.

My loving family, old and new, gave me a lot of strength and support, thank you! I especially thank my parents, Laurence and Matthias, which shaped me into the person I am today and always encouraged me to be ambitious, curious and challenge myself.

Above all, I want to thank Alex for his support and his love. Our timing through our studies and PhD is infamous - being two on the same road makes every challenge easier.